The Cost of Bad Parents: Evidence from the Effects of Parental Incarceration on Children's Education

Carolina Arteaga*

September 24, 2019

[Most recent version]

Abstract

This paper provides evidence that parental incarceration increases children's educational attainment. I collect criminal records for 90,000 low-income parents who have been convicted of a crime in Colombia, which I combine with administrative data on the educational attainment of their children. I exploit exogenous variation in parental incarceration resulting from the random assignment of defendants to judges with different propensities to convict and incarcerate. Because I only observe defendants who are convicted, my identification strategy differs from previous judge IV applications. To address this challenge, I depart from the single dimension threshold setting, and model conviction and incarceration in the context of a multiple dimension threshold crossing model, thereby defining a new policy relevant causal parameter. I find that conditional on conviction, parental incarceration increases education by 0.7 years for children whose parents are on the margin of incarceration.

JEL No. I24,J24,K42

*I am very grateful for the support, guidance, and encouragement of Adriana Lleras-Muney, Rodrigo Pinto, Sarah Reber, and Till von Wachter. I thank Leah Boustan, Moshe Buchinsky, Denis Chetverikov, Christian Dippel, Paola Giuliano, Martin Hackman, Maurizio Mazzocco, Rosa Matzkin, Ricardo Perez-Truglia and Manisha Shah for their feedback. I also thank my colleagues Richard Domurat, Sepehr Ekbatani, Stefano Fiorin, Alex Fon, Keyoung Lee, Rustin Partow, Vitaly Titov, Lucia Yanguas, Diego Zuñiga and seminar participants at Brown, George Washington, Los Andes, Rotman, Rutgers, Syracuse, Toronto, UCLA, Zurich ALCAPONE, Binghamton and CCPR for insightful discussions. Natalia Cardenas, Mauricio Duran, and Juan D. Restrepo provided invaluable help in answering my questions about the institutional context. I gratefully acknowledge support from the Treiman Fellowship, CCPR, Colciencias, and the Central Bank of Colombia. Comments are greatly appreciated. Department of Economics, University of Toronto (carolina.arteaga@utoronto.ca).

1 Introduction

Over one million children in EU countries and 2.7 million children in the U.S. have a parent in prison (Sykes and Pettit, 2014).[1] Family environments during the early years, and especially parenting, are known to be major determinants of human development (Heckman, 2013 and Almond et al., 2019), yet there is only a small literature investigating the effects of parental incarceration on children's outcomes. A large body of correlation-based evidence finds negative associations between parental incarceration and a host of important variables such as mental health, education, and crime (Wakefield, 2015). However, households with incarcerated parents are disadvantaged along many dimensions.[2] Therefore, estimates from differences across outcomes are likely to be negatively biased.

In this paper, I estimate the causal effects of parental incarceration on children's educational attainment in Colombia. To do so, I exploit exogenous variation resulting from the random assignment of defendants to judges with different propensities to convict and incarcerate defendants. I construct a new dataset that links sociodemographic data on households with children from SISBEN, Colombia's census of the low-income population, to criminal records for parents scraped from the internet. I find criminal records for approximately 90,000 parents for the years 2005 to 2016. Then, I link the educational outcomes of criminals' children using administrative data on public school enrollment. Finally I web-scrape the children's criminal records after they turn eighteen years old.

I estimate that on average, conditional on conviction, parental incarceration increases education by 0.7 years for children whose parents were on the margin of going to prison.[3] With an average schooling of 6.8 years, this corresponds to an increase of 10.2%. In addition, the estimated marginal treatment effect (MTE) suggests that the benefit of parental incarceration is larger for children of parents who were incarcerated by more lenient judges. Intuitively, those who are incarcerated by lenient judges have worse unobserved characteristics on average, than those incarcerated by

[1]Sykes and Pettit (2014) also estimate that for the U.S., 62% of black children born to high school dropouts will experience the imprisonment of a parent by age 17.

[2]Even prior to the incarceration event, these households are more likely to be poor and to experience domestic violence (Arditti, 2005; Arditti et al., 2012). In the US, Mumola (2000) finds that 60% of parents in prison reported that they used drugs in the month before their offense, 25% reported a history of alcohol dependence, and about 14% reported a mental illness. Western (2018) also documents that around 60% of parents in prison had experienced childhood trauma, such as domestic violence and sexual abuse.

[3]I refer to those on the margin of going to prison as those whose incarceration sentence would have been different under a harsher or more lenient judge. Given that my instrument is continuous, this estimate is not the effect on a single margin, but the weighted average of all of those for whom judge assignment could have resulted in a different outcome.

the most strict judges. In terms of observed heterogeneity, point estimates suggest that the benefit of parental incarceration is larger when the child is a boy, incarceration was for a violent crime, or the incarcerated parent is the mother, though only the difference in the treatment effects by gender of the child is statistically significant. Lastly, parental incarceration may result in the child being placed with an alternative caregiver who can provide better care to the child. Indeed, I find suggestive evidence which indicates that after an episode of parental incarceration, children often move in with their grandparents. Children are also more likely to move to a household not in SISBEN, which suggests an improvement in economic conditions.

Previous papers in this literature use the random assignment of defendants to judges and their systematic differences in leniency to estimate the causal effects of incarceration on various outcomes.[4] In those settings, the authors observe the universe of defendants who face trial and construct incarceration rates at the judge level.[5] In my setting, I only observe the pool of defendants who were convicted in trial, and who may or may not have been incarcerated. To address this limitation in the data, I model both the selection into conviction, and then the selection into incarceration using a general framework of a multiple dimension threshold model. Treatment can take one of three possible outcomes: i) not convicted, ii) convicted and not incarcerated, and iii) convicted and incarcerated. The crossing the first threshold decides conviction, and for those convicted, crossing the second threshold determines incarceration. In this setting, by looking at the strength of the evidence, judges decide first on conviction, and then, for those convicted, by looking at the severity of the crime, they decide on incarceration. These two distinct groups may have different treatment effects that are of interest to policy makers, and respond to different policy concerns. Conviction is about prosecution and criminal investigation efforts, and incarceration on the other hand, is a matter of punishment or rehabilitation.

I use the technology in Lee and Salanie (2018) and Heckman and Pinto (2018) to identify treatment effects in a setting with multiple threshold-crossing rules.[6] Mine

[4]See Kling (2006); Aizer and Doyle (2015); Di Tella and Schargrodsky (2013); Mueller-Smith (2015); Bhuller et al. (2016); and Dobbie et al. (2018a), among others.

[5]One exception is Mueller-Smith (2015), in his context the data exhibit multidimensional (e.g fines, community service, probation among others) and non-monotonic sentencing patterns and he proposes an estimation procedure that address this challenges. Also, Bhuller et al. (2016) addressed concerns about possible violations of the exclusion restriction given multidimensional sentencing by augmenting the model to include other measures of trial outcomes, and find no evidence of such violation. In my setting, however, the concern is not related to violations of the exclusion or monotonicity assumptions, but to the identification of a parameter of interest.

[6]The model in this paper is not part of the models Lee and Salanie (2018) provide identification

is the first empirical application of a model with multiple dimensional thresholds. Given an instrument for each of the two decision margins, treatment effects related to the second margin (incarceration) can be identified by fixing the crossing of the first threshold, and then exploiting further instrumental variation on the second margin. I estimate a new causal parameter which corresponds to the treatment effect of incarceration that is a function of the the selection into conviction.[7] This weaker identification result is, however, economically relevant: It allows me to estimate the causal effect of incarceration as a function of the selection into conviction. In my empirical exercise, however, I do not find different treatment effects of incarceration along the conviction margin.

The identification result presented in this paper is a general result that applies to any setting that follows the treatment assignment described above. For example, in a context where school admissions are decided upon academic excellence, and for those admitted financial aid is granted on the basis of need, this result provides a way to estimate the causal effect of financial aid for those on a specific level of academic achievement. The effect of financial aid may be different for students who were marginally accepted relative to those with the highest grades, and this provides valuable information for the design of education policy

Contemporaneous to the writing of this article, three papers exploiting judge leniency as an instrument have provided different results using data from Norway, Sweden and Ohio in the US. Dobbie et al. (2019) and Bhuller et al. (2018) find imprecise null effects of parental incarceration on academic achievement for Sweden and Norway, respectively.[8] For Cleveland, Ohio, Norris et al. (2019) find null effects in test scores or grade repetition, but find that parental incarceration causes children to live in higher socio-economic status neighborhoods as adults and decreases the likelihood that a child is incarcerated.

These results are somewhat in contrast to the large positive effects I find for Colombia. Such heterogeneity points to the importance of understanding the settings and the population who identify the treatment effect in each context. Two

results for.

[7]Unconditional treatment effects cannot be identified without further assumptions such as the independence of the two sources of unobserved treatment heterogeneity.

[8]There are many differences between Colombia and Scandinavian countries, some of which may drive these different results. First, the size of the treatment is larger in Colombia, where on average prison sentences are 4.4 years, compared with three and eight months in Sweden and Norway, respectively. A second key difference is the potential size of the effects on schooling before college: In Colombia, 31% of the population between 25 and 34 years old has less than a high school degree, whereas this number is 17% for both Norway and Sweden (OECD, 2016). Finally, Norway and Sweden have very generous welfare programs and better education systems compared with those available in Colombia; these programs help insure disadvantaged children and would also point toward smaller treatment effects in the Scandinavian countries.

key differences can help reconcile these results: first, given the higher incarceration rate in the US, and the lower crime rates in both the Scandinavian countries and the US compared to Colombia, the parents who are incarcerated at the margin in Colombia are more negatively selected than in the US, Norway, or Sweden, in terms of the severity of the crime, but also in terms of income and education. Second, unlike the other papers, my sample consists only of children who lived with their parent prior to the incarceration episode. In the US, half of the parents were not living with their children at the time of incarceration (Parke and Clarke-Stewart, 2002), and as a result the scope for positive effects from removing a parent is more limited. Consistent with this view, other papers that focus on parents living with their children in the US find results similar to mine. Cho (2009) finds that children in Chicago's public schools whose mothers went to prison instead of jail for less than one week are less likely to experience grade retention. Using an event study design, Billings (2018) finds that incarceration improves end-of-grade exams and behavioral outcomes. He also finds, as I do, larger benefits when the mother is the incarcerated parent.

My paper also contributes to the literature examining how parents affect their children's outcomes. This includes a large body of papers on the intergenerational effects of human capital (Black et al., 2005; Oreopoulos et al., 2006), wealth (Black et al., 2015), and welfare receipt (Dahl et al., 2014), among other variables. Specifically, my paper contributes to the literature on household structure and children's outcomes, and shows that living with a parent is not always better for children.[9] Finlay and Neumark (2010) study whether marriage is good for children, and find that unobserved factors drive the negative relationship between never-married motherhood and child education.[10] In addition, there is mixed evidence on the effects of removing children from their parents and placing them in foster care; for South Carolina Roberts (2018) finds positive effects on schooling, Bald et al., (2019) find mixed results across gender and age for Rhode Island, and Doyle (2007, 2008) find negative labor market and crime outcomes for Illinois. The results in my paper suggest that children may benefit from the absence of a convicted parent who is at the margin of incarceration.

The rest of the paper is structured as follows. Section 2 provides background on the judicial system in Colombia, and Section 3 describes the data sources and

[9]Also see Lang and Zagorsky (2001) who find little evidence that a parent's presence during childhood affects economic well being in adulthood.

[10]There is also a literature in sociology on the effects of marital conflict and divorce on children's well-being. Using longitudinal data, Amato et al. (1995) find that in high-conflict families, children have higher levels of well-being as young adults if their parents divorce rather than stay together.

provides summary statistics. Section 4 describes a model to identify causal effects in my setup, Section 5 presents my estimation and results, and Section 6 discusses the results, the mechanism and external validity. Section 7 concludes.

2 Background: The Colombian Court System

In this section, I describe the criminal justice system in Colombia: how defendants are processed, how cases are assigned to judges, the types of crimes involved, and the stages of a standard trial.

Figure 1 illustrates how defendants are processed in Colombia's criminal justice system.[11] A criminal record is created when an arrest is made. Once this happens, the police and a randomly assigned prosecutor must present the evidence that motivated the arrest in front of a judge within 36 hours. This judge, who is randomly assigned from the lowest tier of the judicial hierarchy, determines whether the arrest was legal and whether the defendant should await trial in prison.[12] Next, the case is randomly assigned to another judge who will preside over the trial—this is the judge who provides the exogenous variation in conviction and incarceration I use in this paper. In practice, once the first judge decides to continue with the prosecution of a defendant, the case is entered immediately into a software program that assigns a judge at random among the judges in the judicial district and at the court level that the case is designated to; I refer to the district/court level as the "randomization unit."

Colombia is divided into 33 judicial districts. In the largest cities, a district usually encompasses the city's metropolitan area, and for the rest of the country, it usually corresponds to a state. Depending on the severity of the charge(s), a case will be randomized within one out of three possible court levels within the judicial district in which the crime was committed. The first level, municipal courts, receive simple cases, such as misdemeanors, property crimes involving small amounts, and simple assault cases. These cases account for 38% of the data. More severe crimes, such as violent crimes, drug- or gun-related crimes, and large property crimes are sent to circuit courts (56%). Lastly, the most severe types of crime, such as aggravated homicide or terrorism, are assigned to a specialized judge (6%).[13] On average, there

[11] Acuerdo CSJ, 3329.

[12] A defendant will go to prison before trial when at least one of the following conditions holds: i) the defendant is a danger to society, ii) the defendant can interfere with the judicial investigation, or iii) there is reason to believe that the defendant will not appear in court for trial. Art 308. Criminal Proceedings Code.

[13] Art 35-37, Criminal Proceedings Code.

are 20 judges per randomization unit, and the largest district—Bogota—has 55 judges.

Once the judge is assigned, the prosecutor and defense present their arguments to the judge over the course of multiple hearings. The purpose of the first hearing is to formally press charges. In a second hearing, prosecution and defense present all relevant evidence. Next, based on the strength of the evidence, on a third hearing the judge decides on conviction. If the defendant is found guilty, the judge holds a final hearing to determine sentence length and incarceration considering the severity of the crime, potential future harm to society and any aggravating or mitigating factors. The Colombian Penal Code establishes minimum and maximum sentences for each crime, but there is significant discretion on the part of the judge. The general sentencing guidelines range is often quite broad. For example, prison time for possession of 100 grams of cocaine is between five and nine years (Penal Code, Art 376). The judge also determines the crime and severity of the charge the defendant will ultimately be sentenced for —for example, murder versus involuntary manslaughter.

The decision to send a defendant to prison is determined by the length of the sentence. To deal with prison overcrowding, those convicted only serve time in prison when the sentence is longer than a certain threshold.[14] This threshold is set at the national level and has increased overtime. Currently, a sentence equal to four years or less is not served in prison.[15] As a result, the population that faces a trial is divided into three groups: i) not convicted; ii) convicted and not incarcerated; and iii) convicted and incarcerated. The fact that a portion of the convicted population does not serve time in prison is not a special feature of the Colombian penal system; for example, it is comparable to a sentence of probation in the US.

In Colombia, judges are selected based on their performance on an exam from an open call of attorneys, with specific legal experience requirements for each category of judge. Appointments do not have term limits, and it is common that, over time, judges rise within the judicial hierarchy. The average tenure of a judge is six years, and on average, a judge presides over 344 cases.

While in prison, inmates can receive visits from adults once a week and from their children once a month. The government does not provide special welfare assistance to inmates' families. Unlike in the US, being convicted of a crime does not change

[14]This feature is not unique to the Colombian setting(e.g. Italy) and can also be compared to a probation sentence.

[15]In these cases, the only consequence of being convicted is that for the duration of the sentence, the judge must be notified of any change of address or if the convict plans to travel outside the country. Art 63 Penal Code, and Ley 1709 de 2014.

one's eligibility for welfare benefits, and in the labor market, it is not common practice to ask about previous convictions, although this information is available online.

3 Data Construction

3.1 Data sources

I collect data from several sources. First, I use two waves of Colombia's census of potential beneficiaries of welfare (SISBEN). These data are collected by the government to characterize the country's poor population and to target social programs to them. SISBEN has information on national identification numbers (NINs), household structure, age, gender, education, labor force participation of each household member, and a large set of variables on characteristics and assets of each house (e.g., refrigerator, stove, and floor material, among others). With this information, the government creates a score for each household that summarizes its level of wealth. The score is used to determine eligibility for most public programs—for example, free health insurance, conditional cash transfers, nutrition programs, subsidized housing, and college loans, among many others (Bottia et al., 2012). The first wave, conducted from 2003 to 2005, has data on 31.9 million citizens; the second wave, conducted from 2008 to 2010, has data on 25.6 million citizens.

From this database, I obtain two key elements for my analysis. First, I observe parent and child links when they live in the same household. Second, I use parents' NINs to scrape criminal records. Anecdotal evidence for Colombia suggests that a large share of children with an incarcerated parent were not living with the parent at the time of the crime, all of these cases will not be part of from my sample. My target population is, however, likely to be the most affected by parental incarceration.[16]

In Colombia, criminal records from defendants who are convicted are public and available online for 17 out of 33 judicial districts. These 17 districts represent 67% of the population, 69% of homicides, and 83% of property crimes; they include the largest cities in the country; and they are richer and more urban than the 16 districts without data online.[17] Each criminal record includes the name and NIN

[16] Given how my parent-to-child links are constructed, I focus on parents who are living with the children rather than the biological parents. This definition includes stepchildren when the parent identifies the child as his or her child instead of describing themselves as not being related to the child.

[17] The universe of judicial sentences is public; however, they are only available in the nation's National Archives. Criminal records for Bogotá can be found at the following link: http://procesos.ramajudicial.gov.co/jepms/bogotajepms/conectar.asp

of the defendant, crime, date of crime, sentence information, and the court type and number that handled the case. I collected data on court directories and court identifiers to link each record to a specific judge. There is only one judge per courtroom but judges change over time, I construct the tenure of each judge at each courtroom to assign cases to judges.

I complement these data with individual-level, anonymized records from the Attorney General's Office. This database has information on the universe of criminal cases (including cases that did not result in a conviction), along with courtroom identifiers, date of trial, final verdict, and gender and age of the defendant. I use this information to construct a measure of conviction stringency at the judge level. Finally, I use administrative records of public school enrollment for 2005-2016 with names and NINs to construct a measure of educational attainment. Children's educational attainment is capped at 11, which is the last year of high school in Colombia.

3.2 Sample

To construct my sample, I proceed as follows: From SISBEN, I take the NINs of all parents living with their children in the 17 districts that have information online and web-scrape their criminal records. This adds up to 17 million adults. For computational reasons, I only search for records in the district where the person was living at the time of the SISBEN survey. To assess the number of records I miss due to this restriction, I take a 5% random sample and look for their criminal records in all 17 districts. From this, I estimate that I miss 8.6% of the sample due to crimes committed in districts different from the one found in SISBEN. My sample, therefore, includes only poor parents who, at the time of the SISBEN survey, lived with their children, lived in the largest districts of the country, and committed crimes in the district in which they were living.

I find 328,579 criminal records for 256,108 individuals, of which 63,654 have missing fields in at least one of the key variables, such as court identifier, crime, year, or sentence. Half of these records with missing data correspond to Medellin, which is the second largest district after Bogota, and has missing court identifiers in all of their records. I keep only crimes committed after 2005 and after the year of the first SISBEN year records, which results in 193,520 records.[18] Next, I drop all records from court levels for which there was only one judge (5,963 cases dropped),

[18]In 2005, there was a reform in the judicial system that renders the two periods incomparable. In the previous system, a judge served as both prosecutor and judge at the same time, and he or she was anonymous to the defendant. Additionally, at the time of this reform, there were other changes put in place regarding sentencing guidelines.

and also in cases in which the number of records per judge in a year is fewer than 15 (44,806). I also only keep courtrooms for which I have judge/year conviction rates from the Attorney General's Office database. This leaves me with 128,792 criminal records from 105,133 adults. I retain only the first conviction in my sample, and collect data on the crime, courtroom identifier, and decisions regarding sentence and incarceration.[19] I merge the criminal records back into the SISBEN data and keep only the first parental conviction in the household. My final data set consists of 91,032 convicted parents.

I link these data to two outcome variables for these children: educational attainment and criminal records. I find school records for 77% of them, similar to the share of children between ages 12 and 17 who attend school (76%, 2005 Census). Table C2 in the Appendix shows evidence that having a missing education record is mostly due to actually not being in school, as reported in SISBEN; it is also not statistically related to parental incarceration. I also search for criminal records for all children of convicted parents who were 18 years of age by 2017. My final data set consists of 52,419 children born between 1992 and 2007 who have a convicted parent. In the following section, I characterize the population of convicted and incarcerated individuals, as well as their households and children.

3.3 Summary statistics

The population in my sample is negatively selected along three margins: education, income and criminal activity. In Table 1, I present socioeconomic characteristics for adults in the overall population, for parents in SISBEN with and without a conviction, and for parents with a conviction, by incarceration status. By comparing column 1 and columns 2 and 3, we see that parents in the SISBEN have fewer years of education, are less likely to have a high school degree, live in larger households, and are more likely to be single than all adults. Among parents in the SISBEN, individuals with a conviction are also negatively selected across a host of variables (column 3 relative to column 2). Convicted adults have fewer years of schooling, are less likely to have a high school degree or more (23% vs. 31%), and have lower income scores. They also live in larger households and are more likely to be single (41% vs. 35%, respectively). Adults with criminal records are disproportionately male (84%), they are more likely to work and to be the head of the household than those without a criminal record.[20]

[19] I only keep the first parental conviction to able to assign the child a unique conviction/incarceration and leniency value.

[20] In the US context, for example, 29% of parents in state prisons have a high school degree or more, 48% are single, 92% are male, and the median age is 32 (Mumola, 2000).

Among convicted parents, incarcerated parents have lower education and lower income levels (columns 4 and 5). Gender differences in the probability of incarceration conditional on conviction are far smaller than those in conviction. Incarceration is associated with lower probabilities of working, as well as being the head of the household. Table 2 splits the sample by gender. On average, convicted women have lower levels of education relative to convicted men, and they tend to come from poorer households. Compared to men, women are less likely to be the head of the household; yet they are still much more likely to be the heads of their respective households than in the country's overall female population (36% vs. 29%, respectively). Convicted women are also more likely to be single.

Property crimes are the most common type of offense (25%), followed closely by drug-trafficking crimes (24%). Violent crimes account for 20% of the records, followed by gun-trafficking and misdemeanor offenses at 18% and 12%, respectively. Incarceration rates vary substantially by crime. Figure 2 ranks crimes by their incarceration rates for selected crimes. Serious crimes, such as kidnapping or rape, have the highest incarceration rates, whereas failure to pay child support, simple assault, and property damage have the lowest. In the middle of the distribution, we find crimes such as drug trafficking, domestic violence, counterfeit currency trafficking, theft, and smuggling, among others.

4 Identification

Children from households with incarcerated parents are disadvantaged along many dimensions. As a result, simple comparisons of outcomes from children with and without incarcerated parents would lead to negatively biased estimates of the effects of parental incarceration. A common way to address this endogeneity concern is to exploit the random assignment of defendants to judges who differ in their leniency to incarcerate.[21] The assumption of this identification approach is that selection into incarceration is decided upon the crossing of a threshold rule over a single dimension of unobserved heterogeneity. Judges have different thresholds which creates variation in their leniency, and as a result, for some defendants their incarceration decision will only be determined by whether they were assigned to a harsh or lenient judge.

In this literature, authors have data on the pool of cases randomly assigned across judges and use this to construct their incarceration instrument, as the share incarcerated by each judge —the leave-out mean. In my setting, I only observe

[21] See Kling (2006); Aizer and Doyle (2015); Di Tella and Schargrodsky (2013); Mueller-Smith (2015); and Bhuller et al. (2016), among others.

defendants who are convicted. To address this challenge, I provide a new identification result using the technology in Lee and Salanie (2018) and Heckman and Pinto (2018). Specifically, I consider a multivalued treatment model (No convicted, Convicted and not incarcerated, and convicted and incarcerated), where the selection into conviction and incarceration is determined by the crossing of two distinct thresholds. Section 4.1 presents a simplified framework to provide intuition of the identification result, and next in Section 4.2 I define the model formally.

4.1 A simplified framework

To fix ideas, let us consider the following framework: Judges are randomly assigned to defendants to make conviction and incarceration decisions by evaluating two distinct attributes of the defendant. When deciding on conviction C, a judge assesses the strength of the evidence of the case at hand. Without loss of generality, the distribution of the strength of the evidence across defendants U^c is uniform [0,1], where zero is the smoking gun and one is no evidence against the defendant. The judge can be one of two types in conviction: harsh (H_c) or lenient (L_c). Harsh judges do not require much evidence to convict a defendant. They have a threshold of 0.8, and thus they convict 80% of defendants; this corresponds to all defendants with a level of evidence below 0.8. Lenient judges require more evidence to convict a defendant, choosing a threshold of 0.2, such that they convict only 20%.

Next, if a defendant is convicted, the judge decides on incarceration I. The judge makes this decision based on an assessment of how harmful the convicted defendant may be to society, and how much punishment the defendant deserves. This trait, which I denote U^I, is also distributed uniformly [0,1]. Very harmful defendants have low values of U^I, and non-harmful defendants have values close to 1. Again, regarding incarceration a judge can be either lenient or harsh. A harsh judge (H_I) would send 70% of convicted defendants to prison, whereas a lenient one (L_I) would only incarcerate 30%. It is the same judge making both decisions so a judge can be of one of four types. Figure 3 illustrates this situation. The x-axis traces the strength of the evidence the conviction decision is based on. That is, we can order defendants along one relevant dimension—namely, the strength of the evidence in the [0,1] interval. A judge splits the space into two when she or he sets her or his conviction rate: Defendants to the right are free, and defendants to the left are convicted. Similarly, the y-axis traces the defendant's punishment level, which is related to the assessment of predicted future criminal activity; unobserved—to the econometrician, not the judge—crime severity; and any mitigating/aggravating factors or family ties.[22] I refer to this dimension as a measure of the defendants' overall quality. For a fixed level of evidence required for conviction, a judge's incarceration level splits the space of convicted individuals into two: A defendant below the threshold will go to prison, and a defendant above will not.

Due to randomization, all judges start with a statistically identical pool of defendants. However, after the conviction decision is made, the pool of convicted defendants is no longer comparable across judges with different conviction thresholds. Defendants convicted under a judge who requires solid evidence to convict will have, on average, a stronger case against them than those convicted under a judge who convicts even under weak evidence of guilt.

Defendants convicted under a harsh judge can face two types of judges $[H_c, H_I]$ or $[H_c, L_I]$, where the first term refers to the judge's conviction stringency, and the second refers to the incarceration stringency. Similarly, those convicted under lenient judges can also have judges of types $[L_c, H_I]$ and $[L_c, L_I]$. Within these partitions, defendants are balanced across judges: first, because they were randomly assigned to their judge, and second, because they were selected into conviction under the same threshold. As a result, within partitions, there is exogenous variation in the probability of going to prison. For example, convicted defendants who were assigned to a $[H_c, L_I]$ judge face a 30% chance of incarceration, whereas those assigned to a $[H_c, H_I]$ judge face a 70% probability. Figure 4 illustrates this argument. This means that for 40% of defendants whose harmfulness assessment is located above the worst 30% of the population, but still in the bottom 70%, incarceration is only a function of judge assignment. Thus, I will be able to estimate LATE-type parameters for defendants who fall into this range.

Specifically, for this example I estimate the following two LATE parameters:

$$LATE_{H_c} = E[Y(t_I) - Y(t_c)|U^c < 0.8, 0.3 < U^I < 0.7]$$

and,

$$LATE_{L_c} = E[Y(t_I) - Y(t_c)|U^c < 0.2, 0.3 < U^I < 0.7]$$

Where $LATE_{H_c}$ is the causal effect of incarceration relative to conviction for those convicted under a harsh judge ($U^c < 0.8$), and $LATE_{L_c}$ is the one for conviction under a lenient judge. $Y(t_I)$ and $Y(t_c)$ represent counterfactual outcomes

[22] As mentioned above, sentencing laws guide the judge's incarceration decisions; however, there is large scope for discretion, even within a specific crime. What this dimension tries to capture are the factors that cause a judge to make different incarceration decisions for criminals who have the same charges.

(years of education of the child) for incarceration (I) and conviction (C), and U^c traces the selection on the conviction stage. And,

$$LATE_{H_c} = \frac{E[Y|H_c, H_I] - E[Y|H_c, L_I]}{E[T = I|H_c, H_I] - E[T = I|H_c, L_I]}$$

Where $T = I$ in the denominator represents treatment assignment equal to incarceration. Similarly, we can have the analogous expression for $LATE_{L_c}$.[23]

4.2 Model

In this section, I formalize the previous intuition and extend it to the case of continuous instruments to deliver a new identification result.

The model is described by the standard IV model, which consists of five main random variables: $T, Z, Y, \mathbf{V}, \mathbf{X}$. Those variables lie in the probability space (Ω, F, P), where individuals are represented by elements $i \in \Omega$ of the sample space Ω. The variables are defined below:

- T_i denotes the assigned treatment of individual i, and takes values in $supp(T) = \{t_f, t_c, t_I\}$. t_f stands for not convicted, t_c for convicted but not incarcerated, and t_I for convicted and incarcerated.
- Z_i is the instrumental variable in this analysis and takes values in $supp(Z)$, and represents judge assignment.
- Y_i denotes the outcome of interest for individual i, —e.g., years of education of the child.
- $\mathbf{X}_i$ represents the exogenous characteristics of individual i.
- $\mathbf{V}_i$ stands for the random vector of unobserved characteristics of individual i, and takes values in $supp(\mathbf{V})$.

The random vector $\mathbf{V}$ is the source of selection bias in this model. It causes both the treatment T and outcome Y. The standard IV model is defined by two functions and an independence condition, as follows:

$$\text{Outcome Equation: } Y = f_Y(T, \mathbf{X}, \mathbf{V}, \epsilon_Y) \quad (1)$$

$$\text{Treatment Equation: } T = f_T(Z, \mathbf{X}, \mathbf{V}) \quad (2)$$

$$\text{Independence: } Z \perp (\mathbf{V}, \epsilon_Y)|\mathbf{X} \quad (3)$$

[23] For applications where the universe of trial cases is available you can estimate the additional treatment effect of being convicted.

where ϵ_Y is an unobserved zero-mean error term associated with the outcome equation, that is independent of $\mathbf{V}$.

In this notation, a counterfactual outcome is defined by fixing T to a value $t \in supp(T)$ in the outcome equation. That is, $Y(t) = f_Y(t, \mathbf{V}, \mathbf{X}, \epsilon_Y)$. The observed outcome for individual i is given by:

$$Y = Y(T) = \sum_{t \in \{t_f, t_c, t_I\}} Y(t) \cdot \mathbf{1}[T = t]. \tag{4}$$

The independence condition (3) implies the following exclusion restriction:

$$\text{Exclusion Restriction}: Z \perp Y(t)|\mathbf{X} \text{ for all } t \in supp(T). \tag{5}$$

For the sake of notational simplicity, I suppress exogenous variables $\mathbf{X}$ henceforth. All of the analysis can be understood as conditional on pre-treatment variables.

I assume that the treatment equation is governed by a combination of two threshold-crossing inequalities. First, there is a conviction stage:

$$\begin{cases} \text{Free} & \text{if } \mathbf{1}[\phi_c(\mathbf{V}) > \xi_c(Z)] \\ \text{Convicted} & \text{if } \mathbf{1}[\phi_c(\mathbf{V}) \le \xi_c(Z)]\cdot \end{cases}$$

where $\mathbf{1}[\cdot]$ denotes a binary indicator and $\phi_c(\cdot), \xi_c(\cdot)$ are real-valued functions. Function $\phi_c(\cdot)$ measures the degree of culpability assessed by the judicial system. This function looks at variables and information that are not observed by the econometrician but that are observed by the judge, such as the evidence, crime intensity, effort of the defense and prosecutor lawyers, as well as unobserved characteristics of the defendant such as aggression, antisocial behaviour, etc. The function $\xi_c(\cdot)$ assesses judge leniency on conviction. This function can be understood as a threshold of reasonable doubt beyond which the defendant is convicted by the judge. Judges differ in their leniency and may set different thresholds for evidence. The judge convicts defendant i whenever $\phi_c(V_i) \le \xi_c(Z_j)$. If that is the case, a second stage is held and the judge makes a decision regarding incarceration:

$$\begin{cases} \text{Not incarcerated} & \text{if } \mathbf{1}[\phi_I(\mathbf{V}) > \xi_I(Z)] \\ \text{Incarcerated} & \text{if } \mathbf{1}[\phi_I(\mathbf{V}) \le \xi_I(Z)] \end{cases}$$

Similarly, $\phi_I(\mathbf{V})$ is a function whose arguments are the case and defendant's characteristics relevant for assessment of the punishment level. As before, the judge compares $\phi_I(\mathbf{V})$ to her/his threshold to incarcerate $\xi_I(Z)$.

Treatment assignment can be summarized as follows:[24]

$$T = f_T(Z, \mathbf{V}) = \begin{cases} t_f & \text{if } \mathbf{1}[\phi_c(\mathbf{V}) > \xi_c(Z)] \\ t_c & \text{if } \mathbf{1}[\phi_c(\mathbf{V}) \leq \xi_c(Z)] \cdot \mathbf{1}[\phi_I(\mathbf{V}) > \xi_I(Z)] \\ t_I & \text{if } \mathbf{1}[\phi_c(\mathbf{V}) \leq \xi_c(Z)] \cdot \mathbf{1}[\phi_I(\mathbf{V}) \leq \xi_I(Z)] \end{cases}$$

This model relies on two separable threshold functions that play the role of the monotonicity condition (Vytlacil, 2002).[25] Without loss of generality, it is useful to express treatment assignment using the following variable transformation:

$$U^c = F_{\phi^c(\mathbf{V})}(\phi^c(\mathbf{V})) \sim Unif[0, 1], \tag{6}$$

$$U^I = F_{\phi^I(\mathbf{V})}(\phi^I(\mathbf{V})) \sim Unif[0, 1] \tag{7}$$

where $F_K(\cdot)$ denotes the cumulative distribution function of a random variable K. U^c, U^I are uniformly distributed random variables in $[0, 1]$, and there is no restriction

[24]I assume the following standard regularity conditions: A1) $E(|Y(t)|) < \infty$ for all $t \in supp(T)$, A2) $P(T = t|Z = z) > 0$ for all $t \in supp(T)$ and all $z \in supp(Z)$ and, A3) $(\phi_c(\mathbf{V}), \phi_I(\mathbf{V}))$ are absolutely continuous with respect to Lebesgue measure in $\mathbb{R}^2$. The first assumption guarantees the existence of the expectation, the second one assures that there is a share of the population assigned to each treatment group for every judge, and the third one allows me to apply the Lebesgue differentiation theorem.

[25]Consider two judges, j and j', who see defendants i and i', who differ in their level of culpability. Say i' has more evidence against him than i; namely $\phi_c(i') < \phi_c(i)$. Suppose that judge j convicts defendant i' but not i. Then the threshold function implies that it cannot be the case that judge j' convicts defendant i, but not i'. More generally, let $D_i(j) = \mathbf{1}[T_i(j) = t_c]$ denote the binary indicator that judge j convicts defendant i. Thus if judge j convicts i' but not i, it implies:

$$D_i(j) > D_{i'}(j)$$

Then it cannot be the case that judge j' convicts defendant i, but not i'. Which means:

$$D_i(j) > D_{i'}(j) \rightarrow D_i(j') \geq D_{i'}(j'),$$

which is equivalent to stating that:

$$D_i(j) > D_i(j') \rightarrow D_{i'}(j) \geq D_{i'}(j').$$

We can generalize this to all individuals to arrive at the standard monotonicity assumption of Imbens and Angrist (1994).

on the joint distribution of U^I and U^c.

$$P_c(z) = F_{\phi^c(\mathbf{V})}(\xi^c(Z)); z \in supp(Z), \tag{8}$$

$$P_I(z) = F_{\phi^I(\mathbf{V})}(\xi^I(Z)); z \in supp(Z) \tag{9}$$

Let $P_c(z)$ denote the probability of conviction when $Z = z$. Moreover, independence condition (3) implies $P_c, P_I \perp U^c, U^I$. In this notation, the model can be expressed as:

$$T \equiv f_T(Z, \mathbf{V}) = g_T(U^c, U^I, P_c, P_I) = \begin{cases} t_f & \text{if } \mathbf{1}[U^c > P_c(z)] \\ t_c & \text{if } \mathbf{1}[U^c \leq P_c(z)] \cdot \mathbf{1}[U^I > P_I(z)] \\ t_I & \text{if } \mathbf{1}[U^c \leq P_c(z)] \cdot \mathbf{1}[U^I \leq P_I(z)] \end{cases} \tag{10}$$

In the model, U^c and U^I have the same interpretation as in the previous section, and P_c is interpreted as the share convicted by judge z. Without the assumption of independence of U^c and U^I, variation in incarceration leniency is only identified once I fix the conviction threshold. Thus, the counterfactual of interest are $Y(t_I)$ and $Y(t_c)$ for those who were convicted under $P_c = p_c$. This means that the objective is to identify causal effects of the form: $E(Y(t_I) - Y(t_c)|U^c < p_c)$, which is the same exercise explained in Section 4.1. Let:

$$P_I^*(z) = Pr(U^I < P_I(z)|U^c < P_c(z)) \tag{11}$$

P_I^* is the judge's incarceration probability conditional on conviction.

Proposition*:* *The difference in counterfactual outcomes* $E(Y(t_I) - Y(t_c)|P_I^*(Z), U^c < p_c)$ *is identified from the data as follows*:

$$E(Y(t_I) - Y(t_c)|P_I^*(Z), U^c < p_c) = \tag{12}$$

$$\int_0^1 \frac{\partial E(Y \cdot \mathbf{1}[T \in \{t_c, t_I\}]|P_c(Z) = p_c, P_I^*(Z) = p_I^*)}{\partial p_I^*} dp_I^* \tag{13}$$

See the appendix for the proof.

What this result says is that we can trace the treatment effect of incarceration relative to conviction once we fix a threshold for conviction. We do this by evaluating changes in the outcome variable when we change P_I^*. This delivers the MTE along the unobservable dimension $U^I|U^c < p_c$. The integral over the support of the instrument gives the LATE, or the ATE when the instrument has full support.

The identification result in equation (13) is useful in any settings where treatment assignment follows the design in equation (10). In the context of criminal policy where judges decide on both conviction and incarceration the researcher has two instruments to identify two policy relevant treatment effects. The first one, conviction, takes the form of the traditional LATE in the literature, given that treatment is decided upon the crossing of a single threshold. The second one, the effect of incarceration is only identified as function of the crossing of the first threshold. An example where this result may provide new casual estimates is the following: In a context where school admissions are decided upon academic excellence, and for those admitted financial aid is granted on the basis of need, the result in (13) provides a way to estimate the causal effect of financial aid for those on a specific level of academic achievement. The effect of financial aid may be different for students who were marginally accepted relative to those with the highest grades, and this provides valuable information for the design of education policy.

5 Estimation

To apply the identification result of the previous section, I start by estimating the sample analogs of the conviction ($P_c(Z)$) and incarceration ($P_I^*(Z)$) instruments in the model. The interpretation of these variables is the probability of being convicted/incarcerated, given the assignment to a specific judge. Following the literature, these are estimated as judge fixed effects from regressions after parsing out variation at the unit at which the randomization of judges occurred and specific case characteristics. That is, the conviction/incarceration decision can be decomposed into a portion that is related to the individual, the judge, the offense, and the randomization unit/year. I do this as follows:

$$D_{itorz} = \gamma_{rt} + \gamma_o + \epsilon_{itorz}$$

Where D_{itorz} corresponds to a conviction or incarceration dummy, i indexes individuals, t year, o offense, r court-level/judicial district and z judge. γ_{rt} corresponds to randomization-level fixed effects, which is a court-level/judicial-district by year-level fixed effect. γ_o is a offense-level fixed effect (161 different crimes); and ϵ_{itorz} is a mean zero term. Following the literature, I estimate the judge instrument $\widehat{p_{z-i}}$ for defendant i to be the following leave-out estimator:

$$\widehat{p_{z-i}} = \frac{1}{n_z - 1} \sum_{k \neq i} \widehat{res_{z,k}}$$

where n_z is the number of cases of judge z, and res_{zk} is the residual from a regression of the conviction/incarceration dummy on γ_{rt} and γ_o.

Figure 5 shows the distribution of conviction and incarceration rates at the judge level, and $\widehat{p_z}$ for both conviction and incarceration. From the graph, we can see that although court-level/year and crime-level fixed effects explain most of the variation, judge's fixed effects still represent a sizable share of the variance in conviction and incarceration.

5.1 Instrument validity

Next, I examine how much judge fixed effects predict individual-level decisions by estimating a first-stage regression, as follows:

$$D_{itorz} = \beta_0 + \widehat{p_{zi}} + \beta_1 X_i + \epsilon_{itorz}$$

As before, D_{itorz} corresponds to the conviction or incarceration dummy, and p_z is the leave-out mean of judge z assigned to person i. I run this regression with and without controls X_i. In the conviction regression, where I use anonymized data from the Attorney General's Office, I can only control for age, gender, and number of crimes charged.[26] In the incarceration regression, I control for schooling, income, occupation, gender, year of birth, and year in the survey. According to the results in Table 3, judges have a strong influence on conviction and incarceration decisions. The estimates are highly significant and suggest that being assigned to a judge with a 10 percentage point higher conviction/incarceration rate increases the defendant's probability of conviction and incarceration by seven and eight percentage points, respectively. This relationship is robust to the inclusion of controls, as expected by random assignment. Figure 6 depicts this first-stage relationship for conviction (left panel) and incarceration (right panel). These graphs show a strong positive relationship between the instrument and individual trial decisions. The F-stats on the first stage correspond to regressions on judge dummy variables to account for the true dimensionality of the instruments. These F-stats are above the critical value for the leave-out mean instrument for weak instruments (see Figure 4 in Stock et al., 2002). See Section 5.4 for a further discussion about the F-stat.

Recall from the previous section that the variation in incarceration stringency conditional on a level of conviction stringency is what identifies treatment effects in this context. Figure 7 shows a scatter plot of both conviction and incarceration

[26] This extra case variables are included in the system at the discretion of the (randomly assigned) prosecutor and are missing on a considerable share of the cases.

fixed effects. From the graph we can see that there is substantial variation along the incarceration axis for each conviction rate.

For the instrument to be valid, the judge's fixed effects must be orthogonal to the defendant's characteristics. I test this in the anonymized data from the Attorney General's Office, where the universe of cases the judge has heard is available. Table 4 checks the balance across defendants for my judge-stringency measures for conviction and incarceration. Across gender, age, and type of crime—which are the only variables available in these data—I find no individual or joint statistical significance. In addition, the identification result is supported by the observation that once P_c is fixed, the pool of convicted defendants is balanced across judges. I test whether covariates are associated with incarceration stringency for the convicted sample, once I split the sample by conviction group (low, medium, or high) or control for the conviction level with a polynomial of P_c. In Table 5, I test the individual and joint significance of variables associated with education, income, and occupation status, and find no evidence of a relationship with judge stringency.

To interpret the results of the IV as the causal effect of incarceration, judge stringency must only affect child's outcomes through incarceration. This may not be the case if the judge fixed effects capture other dimensions of trial decisions, such as fines or guilt (Mueller-Smith, 2017). In my setting, this is less of a concern because in the case of Colombia, fines are rare and only associated with large property crimes; and because I model the conviction decision directly.[27]

Finally, I also require that conviction or incarceration decisions made by a lenient judge would also have been made by a stricter judge; this is called the monotonicity assumption. One testable implication of monotonicity is that first-stage estimates should be non-negative for all sub-samples (Bhuller et al, 2016). That is for example if a judge is lenient, he or she is going to be lenient for both women and men, and for both violent crimes and nonviolent crimes. To test this assumption, I construct judge fixed effects for just one group in the population, for example, for men and use this fixed effect in a first-stage regression to predict individual conviction and incarceration for women. I do this for gender, type of crime, and age group. Table 6 shows these first-stage tests, in which I find positive first-stage estimates across all slices of the data, which supports the monotonicity assumption. However, if only these weaker monotonicity assumptions holds inference is constrained. In particular, it does not allow for the identification of marginal effects along the entire distribution of judge propensities, as can be achieved in the conventional framework. The weaker

[27]In addition, the failure to pay these fines does not entail any consequence in terms of incarceration.

assumptions rely on averaging across the entire set of judges, while identification of marginal effects throughout the distribution requires assumptions to hold judge by judge (Norris, 2019). In Table 7 I test pairwise monotonicity following Norris (2019) and find I can not reject monotonicity across individuals characteristics, and it is only rejected for property vs not property crimes.[28] Frandsen et al (2019) show that under the usual assumptions, average outcomes by judge will be a continuous function with bounded slope of judge propensities to incarcerate. Intuitively, if this is not the case, it implies that either judges influence outcomes beyond their propensity to assign treatment, or judges disagree on their implicit ordering of which defendants should be treated. Based on that result, they develop a test that jointly test violations to the monotonicity assumption and the exclusion restriction. In Table 8 I implement their test and I find there is no evidence of violation of these assumptions. Finally, I also implement the procedure developed by Mourifié and Wan (2017) to test jointly the monotonicity and independence assumptions, and find that the null hypothesis is not rejected.[29]

5.2 Results

Following the identification result, I need to account for the different levels of conviction stringency at which defendants were found guilty. I do this in two ways: First, I sort my data by stringency in the conviction stage (P_c) and split the sample into terciles: low ($0.7 < P_c < 0.88$), medium ($0.88 < P_c < 0.9$), and high ($0.9 < P_c < 1$) conviction levels. Second, I pool the data and add a second-degree polynomial on P_c with interaction terms. This last estimate can be interpreted as an average effect across the different conviction thresholds. The first column of Tables 9 and 10 show the pooled regression, and the following three columns show the regressions for the split sample.

I begin by showing the OLS estimate of this design. Table 9 shows a regression of parental incarceration on years of education. Following Abadie et al. (2017), I cluster standard errors at the randomization level. Without controls, a child whose parent went to prison has around 0.3 fewer years of schooling than a child whose parent did not. Once I add controls, this difference reduces drastically to less than 0.1 years. Still, we expect that incarcerated parents are negatively selected on unobservables that cannot be accounted for, so -0.1 years is a lower bound on the causal effect.

[28] I split judge leniency across this characteristic and find very similar point estimates.

[29] I implement this test for the following groups: Convicted father, mothers, girls, boys and by crime category.

Next, Figure 8 shows a graphical representation of the reduced-form regression. This graph plots the distribution of judges' incarceration fixed effects against the predicted years of education from a local polynomial regression. From the graph, we can see that there is a strong positive relationship between judge stringency in incarceration and years of education. That is, as we move to the right, where the probability of having a parent in prison increases exogenously, I estimate that the years of education also increase. The top panel of Table 10 shows the regression results for this reduced form: I estimate large increases in years of education for all specifications the increase in years of education is statistically significant. Finally, the bottom panel of Table 10 shows results from the IV; I estimate that having an incarcerated parent increases years of schooling by around 0.7 years on average for all convictions levels. These estimates are statistically different from zero. I find that the increase in years of education is mostly accrued through higher graduation rate from middle school. Figure D1 in the Appendix plots the treatment effect of parental incarceration on grade completed from 6th grade to 11th grade. There are positive treatment effects for all grades, but the effect is larger for 9th grade which corresponds to the last grade of middle school.

I also study how parental incarceration affects the chance that the child is later convicted of a crime. For this exercise, I restrict the data to children who were 18 years old by 2017, so that their criminal records would be public. Figure D2 graphically depicts reduced-form estimates of judge stringency on conviction probability; the effect is close to zero. However, the analysis is under-powered to detect to estimate reasonably sized treatment effects. This is not surprising, since conviction is a low incidence event; only 1.6% of children had a criminal record, and the difference in the OLS is only 0.1 pp.

5.3 Heterogeneity

I this section I examine the heterogeneity of the results along observables and unobservables. In my context, marginal treatment effects (MTE) are particularly interesting, because they trace the causal effect of incarceration along parents' unobserved characteristics (U^I) that matter for incarceration and that are correlated with defendants' quality, broadly defined. What this exercise does is to evaluate the possibility of different effects of parental incarceration given the type of defendant that is going to prison, which is characterized by his or her location along the y-axis of Figure 3. The intuition is as follows: Parents who are incarcerated under the most lenient judges have worse characteristics than those incarcerated under strict judges. In other words, a strict judge incarcerates almost everyone, but a lenient judge incarcerates only the worst defendants, so that those incarcerated under relatively lenient judges are more negatively selected.[30] I follow Heckman and Vitlacyl (2005) to estimate this MTE. Under the assumption of pair-wise monotonicity, I find that at the 5% level, there are heterogeneous treatment effects along parental quality (Figure 9). Specifically, I find that the positive effects of incarceration on schooling accrue when the worst defendants go to prison.

The magnitude of the effect of parental incarceration on children's education is a function of the relationship between the parent and the child prior to the incarceration episode, the type or quality of this parent, and the role of the child in the household. To document this heterogeneity, I estimate the IV regression for different subgroups in the data. Following previous literature in economics, as well as that in psychology and sociology, I estimate different regressions by gender of the child, gender of the parent, birth order, and the nature of the offense—violent, property, drug- or gun-related, and misdemeanor. In Table 11 I show IV results for the pooled model for these different groups in the data.

According to the estimates, the benefits of parental incarceration are larger for boys than girls, and this difference is statistically significant. Specifically, I find that boys' schooling increases by 0.86 years, whereas girls' schooling increases by 0.36 years. This result is consistent with previous research in psychology and economics, which documents that boys are more vulnerable than girls to negative experiences in the household (Bertrand & Pan (2013); Autor et al. (2016); Parke & Clarke-Stewart (2002); Hetherington et al., 1998). Specifically, Autor et al. find that boys, relative to their sisters, have higher rates of disciplinary problems, lower achievement scores, and fewer high school completions when growing up in disadvantaged environments.

I split the sample by gender of the parent and find that incarceration is more beneficial in cases in which the mother is the one going to prison. This result might be surprising at first glance. However, it is important to bear in mind that children's well-being is more closely affected by their mothers' behavior because of their main role as primary caregivers, and that criminal women are more negatively selected than criminal men (Table 2). This result is consistent with the findings of previous research in the US, where Billings (2018) and Turanovic et al. (2012) estimate larger positive effects from maternal incarceration.[31]

[30] I look at this empirically and find that among incarcerated defendants, those incarcerated under stricter judges tend to have fewer and less severe charges. This follows almost directly from the definition of leniency, but also helps to illustrate the way in which these defendants are better.

[31] It is also the case that in the US, incarcerated women have worse socioeconomic backgrounds than incarcerated men (Harrison & Beck, 2006). In addition, Glaze and Maruschak (2008) survey incarcerated parents and find that 60% of imprisoned mothers, compared to 16% of fathers, have histories of being physically or sexually abused.

A source of heterogeneity associated with the quality of the parent going to prison is the type of crime they committed. Thus, in the lower panel of Table 11 I split the sample by crime categories: violent, property, drug-related and gun-related. The largest benefits are observed for defendants convicted for violent crimes, whereas the smaller benefits are for property crimes. These differences, however, are not statistically significant. Nonetheless, this is in line with the previous result on unobserved heterogeneity, in which the positive effects are a function of how good the defendant is as a parent.

5.4 Robustness

In this section I go over various exercises that evaluate the robustness of the results in the paper along different dimensions.

In Table 3 I report the first-stage regression on incarceration, and in the bottom of the table I report the F-test on the excluded instruments. This F-test corrects for the fact that the dimensionality of the instrument is the number of judges and not one (my measure of judge leniency). With this correction, the F-stats are low, but above the critical values for weak instruments. The consequence of weak instruments is that 2SLS-IV estimates will be biased toward the OLS (Stock et al., 2002). In my context, given that the OLS estimates are negative, the bias of the OLS is also negative, and the 2SLS IV estimates are positive, this means that we could expect even larger positive effects. To assess the size of this residual bias, I estimate the IV using the LIML estimator, which is less sensitive to weak instruments—the bias does not increase with the number of instruments (Rothenberg, 1993; Stock et al., 2002). Table D3 in the Appendix shows the estimates for the LIML estimator. I find that the 2SLS and LIML estimator are very close and both are around a point estimate of 0.8 years.

In the Results section, I show my preferred specifications for the estimates of the effect of parental incarceration on educational attainment. This decision to split the sample into three groups of P_c was arbitrary. To assess the robustness of the results, in Figure D3 I instead order observations along P_c, and run multiple regressions on a rolling window of 18,000 observations over P_c, moving the window 500 observations each time. Figure D3 in the Appendix shows that for each sample, I find a positive effect of incarceration on education.

Lastly, as a placebo check, I evaluate whether there are differences in schooling for children of incarcerated versus non-incarcerated parents before the date of the sentence. Table D3 in the Appendix shows that there is no supporting evidence that the positive effects I estimate are the result of preexisting differences in educational attainment.

6 Mechanisms

6.1 What explains the positive effect?

The results presented here suggest that living with a convicted parent has negative consequences for their children. There are many reasons to believe that this is plausible. First, criminals are more likely to exert psychological and physical violence at home, and this can often be detrimental to a child's well-being. In the US context, Western et al. (2004) find that incarcerated men engage in domestic violence at a rate about four times higher than the rest of the population. Furthermore, research in psychology documents that spending time with parents who engage in high levels of antisocial behavior is associated with more conduct problems for their children (Jaffee et al., 2003). This literature concludes that the salutary effects of being raised by married biological parents depend on the quality of care the parents provide.

Second, Chimeli and Soares (2017) document the causal effect of trading illegal commodities on violence. In light of their work, we can expect that households that take part in illegal businesses face constant violence or threats of violence related to guaranteeing property rights or resolving disputes within the business, all of which affect the quality of life in a household. There is also literature on the intergenerational transmission of violence, substance abuse, and crime. Specifically, in the role-model theory, in which children directly observe and model their parents' behavior, incarcerating parents could be beneficial, as it removes bad role models from the house and forces children to update their beliefs about the consequences of criminal behavior (Hjalmarsson and Lindquist, 2012). Beyond intergenerational transmission, childhood exposure to negative behaviors is documented to have direct adverse effects on outcomes in both childhood and adulthood (Balsa, 2008; Chatterji and Markowitz, 2000), all of which helps explain the positive estimates in this paper.

6.2 How does the environment of the child change?

To characterize the changes that households and children experience after an episode of incarceration, I analyze households for which I have two observations in the SISBEN (44% of cases), in which the parent was convicted of a crime between observations. Appearing in both waves of the SISBEN is not random; on the contrary, leaving the sample is associated with an improvement in living standards. This is particularly relevant for children who might be moving to a household outside of SISBEN after the episode of parental incarceration. With this caveat, Table 12 shows suggestive evidence that incarceration is associated with an increase in labor force participation (LFP) of the spouse, a worsening of the income score of the household, and a decrease in the probability of a male as the head of the household. I also find that the probability of living with grandparents increases and the probability of being in the second wave of SISBEN decreases, suggesting that incarceration induces children to move in with relatives who are better off financially.

6.3 Parents at the margin

To derive policy implications, it is important to acknowledge the local feature of my estimates. This paper estimates the effects of parental incarceration for a particular sub-population: children of convicted poor parents at the margin of incarceration. A large share of those convicted—for example, those guilty of murder or rape—would be incarcerated regardless of judge assignment, and this paper cannot provide any insights into the effects on educational attainment of the children of those individuals. At the other end of the distribution, defendants convicted of minor crimes will also avoid prison, regardless of judge assignment. Defendants convicted of drug- or gun-trafficking, domestic violence, and medium-sized property crimes compose the complier group in my estimation, and they are the group my estimates apply to. This marginal population, however, is particularly relevant because it is the population that is more likely to be affected by policy interventions to the criminal justice system. Following Dahl et al. (2014), I find that compliers make up approximately 29.8% of the sample.[32]

6.4 External validity and policy implications

To assess the external validity of my results, I provide a framework motivated by my heterogeneity analysis, which links parental quality to parenting treatment effects

[32] Parental compliers are defendants who would have received a different incarceration decision had their case been assigned to the most lenient judge instead of the strictest judge. We can define the size of this group (π_c) as follows:

$$\pi_c = Prob(Incarceration = 1|z_j = \bar{z}) - Prob(Incarceration = 1|z_j = \underline{z})$$

where $\bar{z}$ and $\underline{z}$ correspond to the incarceration rates of a judge at the 99th and 1st percentiles, respectively. Because of monotonicity, the share of parents who would go to prison regardless of the judge assigned to their case—always takers—is given by the incarceration rate for the most lenient judge and is equal to 22.5%. On the other hand, 47.7% of the sample are children of never takers who would not go to prison no matter which judge was assigned to their case. I estimate that children of compliers make up approximately 29.8% of the sample.

and the probability of incarceration. Figure 10 summarizes this framework. The x-axis traces parental quality; as we move to the right, parental quality increases. The y-axis measures the treatment effect of parenting: Having better parents is better for children. Most importantly, however, there is a segment in the support of parental quality for which parents are detrimental for children. The secondary y-axis measures incarceration probability: In the model, the probability of being incarcerated decreases when parental quality increases. Each society chooses a level of incarceration, which is characterized by a threshold in the support of parental quality. This threshold determines the average effect of incarcerating parents (the gray area in Figure 10). To determine how much the results in this paper apply to other settings, we need to think about the location of the incarceration threshold along the parental quality axis and the shape of the function of the treatment effects' of parents in each country. Countries with higher incarceration rates will incarcerate, on average, better parents than those with lower rates, and as a result we should expect lower benefits or even costs from parental incarceration. We can also expect a much flatter function of treatment effects of parenting in generous welfare states, such as the Nordic countries, in which children's education and health vary less with parental characteristics. As a consequence, we would find smaller treatment effects of parental incarceration (both positive and negative). Similarly, some of the estimates in the literature (Norris et al 2019 and Dobbie et al 2019) consider birth parents who may not necessarily co-reside with their children, in this framework we can hypothesized that it translates to smaller treatment effect of parents and as a result into a smaller effect of parental incarceration.

7 Conclusions

The rise in incarceration has led to an increase in the number of children growing up with a parent in prison. In this paper, I estimate the causal effects of parental incarceration on educational attainment in Colombia. My results suggest that children benefit when their convicted parents are incarcerated. Specifically, I estimate that parental incarceration increases schooling by 0.7 years on average.

In this paper I model both the selection into conviction, and then the selection into incarceration using a general framework of a multiple dimension threshold model. I estimate a new causal parameter which corresponds to the treatment effect of incarceration that is a function of the the selection into conviction. This is the first empirical application of a model with multiple dimensional thresholds. Future possible applications of this identification result could include the identification of treatment effects of financial aid as a function of students academic credentials.

I conclude with a discussion of three important limitations of this paper. First, I consider only the short-term effects of parental incarceration. This is important, as these parents eventually leave prison and will perhaps return to live with their children. Further, if incarceration decreases one's human capital and social and emotional skills, the type of parent who returns after incarceration can be much worse than the one who left. In that case, the long-term effects may be very different from what I estimate here. Another significant limitation of this paper is that, effectively, I can only study one outcome variable. As shown by Dobbie et al. (2018), parental incarceration can have sizable effects on other variables such as earnings and teen pregnancy. These are important results that help characterize the complex shock of having an incarcerated parent, but due to data limitations, I cannot explore them here. Finally, my paper only offers suggestive evidence on the mechanisms that explain the positive effects of parental incarceration on children's educational attainment, further research is required to characterize the obstacles children face in these households in order to provide informed policy recommendations.

References

Abadie, A., Athey, S., Imbens, G.W. & Wooldridge, J., (2017). When Should You Adjust Standard Errors for Clustering? (No. w24003). National Bureau of Economic Research.

Abadie, A. (2003). Semiparametric Instrumental Variable Estimation of Treatment Response Models," Journal of Econometrics, 113(2), 231263.

Aizer, A. & J. J. Doyle (2015). Juvenile Incarceration, Human Capital and Future Crime: Evidence from Randomly-Assigned Judges. The Quarterly Journal of Economics 130 (2), 759–803.

Amato, P. R., Loomis, L. S., & Booth, A. (1995). Parental divorce, marital conflict, and offspring well-being during early adulthood. Social Forces, 73(3), 895-915.

Arditti, J.A., 2015. Family process perspective on the heterogeneous effects of maternal incarceration on child wellbeing. Criminology and Public Policy, 14(1), pp.169-182.

Arditti, Joyce (2012). Parental incarceration and the family: Psychological and social effects of imprisonment on children, parents, and caregivers. New York, NY: New York University Press.

Arditti, Joyce, Sara A. Smock, & Tiffaney S. Parkman (2005). It's been hard to be a father: A qualitative exploration of incarcerated fatherhood. Fathering 3:267–83.

Autor, D., Figlio, D., Karbownik, K., Roth, J., & Wasserman, M. (2016). Family disadvantage and the gender gap in behavioral and educational outcomes (No. w22267). National Bureau of Economic Research.

Bald, A., Chyn, E., Hastings, J. S., and Machelett, M. (2019). The Causal Impact of Removing Children from Abusive and Neglectful Homes (No. w25419). National Bureau of Economic Research.

Balsa, A. I. (2008). Parental problem-drinking and adult children's labor market outcomes. Journal of Human Resources, 43(2), 454-486.

Bertrand, M. & Pan, J., 2013. The trouble with boys: Social influences and the gender gap in disruptive behavior. American Economic Journal: Applied Economics, 5(1), pp.32-64.

Bhuller, Manudeep, Gordon B. Dahl, Katrine V. Løken, & Magne Mogstad. 2017. "Incarceration, Recidivism and Employment." Quarterly Journal of Economics. 22648.

Bhuller, M., Dahl, G. B., Loken, K. V., & Mogstad, M. (2018). Intergenerational effects of incarceration. In AEA Papers and Proceedings (Vol. 108, pp. 234-40).

Billings, Stephen (2017) Parental Arrest and Incarceration: How Does it Impact the Children? (Preliminary draft)

Black, S.E., Devereux, P.J. & Salvanes, K.G., 2005. Why the apple doesn't fall far: Understanding intergenerational transmission of human capital. American Economic Review, 95(1), pp.437-449.

Black, S. E., Devereux, P. J., & Salvanes, K. G. (2005). The more the merrier? The effect of family size and birth order on children's education. The Quarterly Journal of Economics, 120(2), 669-700.

Chimeli, A. B., and Soares, R. R. (2017). The use of violence in illegal markets: Evidence from mahogany trade in the Brazilian Amazon. American Economic Journal: Applied Economics, 9(4), 30-57.

Cho, Rosa M. 2009a. "The Impact of Maternal Imprisonment on Children's Probability of Grade Retention: Results from Chicago Public Schools." Journal of Urban Economics, 65(1): 11-23.

Cho, Rosa M. 2009b. "The Impact of Maternal Incarceration on Children's Educational Achievement: Results from Chicago Public Schools." Journal of Human Resources, 44(3): 772-797.

Criminal Proceeding Code (2004). Codigo de Procedimiento Penal. Ley 906 de 2004; Bogota, Colombia.

Cunha, F., I. T. Elo, and J. Culhane (2013). Eliciting maternal expectations about the technology of cognitive skill formation. Working Paper 19144, NBER.

Currie, J. and Moretti, E., 2003. Mother's education and the intergenerational transmission of human capital: Evidence from college openings. The Quarterly Journal of Economics, 118(4), pp.1495-1532.

Dahl, G. B., A. R. Kostøl, and M. Mogstad (2014). Family Welfare Cultures. The Quarterly Journal of Economics 129 (4), 1711–1752.

Di Tella, R. and E. Schargrodsky (2013). Criminal Recidivism after Prison and Electronic Monitoring. Journal of Political Economy 121 (1), 28–73.

Dobbie, W., Goldin, J., & Yang, C. S. (2018). The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges. American Economic Review, 108(2), 201-40.

Dobbie, W., H. Grönqvistz, S. Niknami, M. Palme and M. Priksk (2018). The Intergenerational Effects of Parental Incarceration. NBER Working Paper, January.

Doyle Jr, J. J. (2007). Child protection and child outcomes: Measuring the effects of foster care. American Economic Review, 97(5), 1583-1610.

Doyle Jr, J. J. (2008). "Child Protection and Adult Crime: Using Investigator Assignment to Estimate Causal Effects of Foster Care." Journal of Political Economy, 116(4): 746-770.

Ehrensaft, M.K., Cohen, P., Brown, J., Smailes, E., Chen, H. and Johnson, J.G., 2003. Intergenerational transmission of partner violence: a 20-year prospective study. Journal of consulting and clinical psychology, 71(4), p.741.

Finlay, K., and Neumark, D. (2010). Is marriage always good for children? Evidence from families affected by incarceration. Journal of Human Resources, 45(4), 1046-1088.

Frandsen, B. R., Lefgren, L. J., and Leslie, E. C. (2019). Judging Judge Fixed Effects (No. w25528). National Bureau of Economic Research.

Furstenberg, F. F., Jr. (1995). Fathering in the inner city: Paternal participation and public policy. In W. Marsiglio (Ed.), Research on men and masculinities series, 7. Fatherhood: Contemporary theory, research, and social policy (pp. 119-147). Thousand Oaks, CA, US: Sage Publications, Inc.

Fomin, S. V. (1999). Elements of the theory of functions and functional analysis (Vol. 1). Courier Corporation.

Imbens, G.W., and D. B. Rubin (1997). Estimating Outcome Distributions for Compliers in Instrumental Variables Models," The Review of Economic Studies, 64(4).

Harrison, P. M. and A. J. Beck (2006). Prison and jail inmates at midyear 2005.

Hart, B. and T. R. Risley (1995). Meaningful differences in the everyday experience of young American children. Baltimore, MD: P.H. Brookes.

Heckman, James J., and Rodrigo Pinto. "Unordered monotonicity." Econometrica 86, no. 1 (2018): 1-35.

Heckman, J. J., Urzua, S., and Vytlacil, E. (2006). Understanding instrumental variables in models with essential heterogeneity. The Review of Economics and Statistics, 88(3), 389-432.

Heckman, James J., and Edward Vytlacil. 2005. "Structural Equations, Treatment Effects, and Econometric Policy Evaluation." Econometrica, 73(3): 669-738.

Heckman, J. J., S. H. Moon, R. Pinto, P. A. Savelyev, and A. Q. Yavitz (2010a). Analyzing social experiments as implemented: A reexamination of the evidence from the HighScope Perry Preschool Program. Quantitative Economics 1 (1), 1–46.

Heckman, J. J., S. H. Moon, R. Pinto, P. A. Savelyev, and A. Q. Yavitz (2010b). The rate of return to the HighScope Perry Preschool Program. Journal of Public Economics 94 (1–2), 114–128.

Heckman, J. J. (2013). Giving kids a fair chance. Mit Press.

Hetherington, E. M., Bridges, M., and Insabella, G. M. (1998). What matters? What does not? Five perspectives on the association between marital transitions and children's adjustment. American Psychologist, 53(2), 167.

Hjalmarsson, Randi, and Matthew J. Lindquist. 2011. "The Origins of Intergenerational Associations in Crime: Lessons from Swedish Adoption Data." Labour Economics, 20: 68-81.

Hjalmarsson, Randi, and Matthew J. Lindquist. 2012. "Like Godfather, Like Son: Exploring the Intergenerational Nature of Crime." Journal of Human Resources, 47(2): 550-582.

Hjalmarsson, Randi, Helena Holmlund, and Matthew J. Lindquist. 2015. "The Effect of Education on Criminal Convictions and Incarceration: Causal Evidence from Micro-data." Economic Journal, 125(587): 1290-1326.

Jaffee, S. R., Moffitt, T. E., Caspi, A., and Taylor, A. (2003). Life with (or without) father: The benefits of living with two biological parents depend on the father's antisocial behavior. Child development, 74(1), 109-126.

Johnson, R. (2009). Ever-increasing levels of parental incarceration and the consequences for children. Do prisons make us safer? The benefits and costs of the prison boom, 177-206.

Kalil, A., 2015. Inequality begins at home: The role of parenting in the diverging destinies of rich and poor children. In Families in an era of increasing inequality (pp. 63-82). Springer, Cham.

Kim-Cohen, J., Moffitt, T.E., Taylor, A., Pawlby, S.J. and Caspi, A., 2005. Maternal depression and children's antisocial behavior: nature and nurture effects. JAMA Archives of general psychiatry, 62(2), pp.173-181.

Kling, J. R. (2006). Incarceration Length, Employment, and Earnings. The American Economic Review 96 (3), 863–876.

Lang, K., and Zagorsky, J. L. (2001). Does growing up with a parent absent really hurt?. Journal of human Resources, 253-273.

Lee, S., and Salanié, B. (2018). Identifying effects of multivalued treatments. Econometrica, 86(6), 1939-1963.

Lefgren, L., Sims, D. and Lindquist, M.J., 2012. Rich dad, smart dad: Decomposing the intergenerational transmission of income. Journal of Political Economy, 120(2), pp.268-303.

Lyle, D. S. (2006). Using military deployments and job assignments to estimate the effect of parental absences and household relocations on children's academic achievement. Journal of Labor Economics, 24(2), 319-350.

McLanahan, S., Tach, L., and Schneider, D. (2013). The causal effects of father absence. Annual review of sociology, 39, 399-427.

Mourifié, I., and Wan, Y. (2017). Testing local average treatment effect assumptions. Review of Economics and Statistics, 99(2), 305-313.

Mueller-Smith, M. (2017). The Criminal and Labor Market Impacts of Incarceration. University of Michigan Working Paper.

Murray, Joseph, and David P. Farrington. 2005. “Parental Imprisonment: Effects on Boys' Antisocial Behaviour and Delinquency Through the Life-Course.” Journal of Child Psychology and Psychiatry, 46(12): 1269-1278.

Murray, Joseph, David P. Farrington, and Ivana Sekol. 2012. “Children's Antisocial Behavior, Mental Health, Drug Use, and Educational Performance After Parental Incarceration: A Systematic Review and Meta-analysis.” Psychological Bulletin, 138(2): 175-210.

Murray, Joseph, Rolf Loeber, and Dustin Pardini. 2012. “Parental Involvement in the Criminal Justice System and the Development of Youth Theft Marijuana Use, Depression and Poor Academic Performance.” Criminology, 50(1): 255-302.

Murray, Joseph, Carl-Gunnar Janson, and David P. Farrington. 2007. “Crime in Adult Offspring of Prisoners: A Cross-National Comparison of Two Longitudinal Samples.” Criminal Justice and Behavior, 34(1): 133-149.

Norris, S. (2018). Judicial Errors: Evidence from Refugee Appeals. University of Chicago, Becker Friedman Institute for Economics Working Paper, (2018-75).

Parke, R. D. and Clarke-Stewart, K. A. (2003). The effects of parental incarceration on children, Prisoners once removed: The impact of incarceration and reentry on children, families, and communities pp. 189–232.

Norris, S., Pecenco, M. and Weaver, J. (2018). "The Effects of Parental and Sibling Incarceration: Evidence from Ohio". Working paper

Pew Center. 2011. “State of Recidivism: The Revolving Door of America's Prisons.” Washington, DC: Pew Charitable Trusts.

Kelsey Roberts. 2018. "Fostering Better Educational Outcomes in Youth" Mimeo.

Stith, S.M., Rosen, K.H., Middleton, K.A., Busch, A.L., Lundeberg, K. and Carlton, R.P., 2000. The intergenerational transmission of spouse abuse: A meta-analysis. Journal of Marriage and Family, 62(3), pp.640-654.

Stock, J. H., Wright, J. H., & Yogo, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. Journal of Business & Economic Statistics, 20(4), 518-529.

Turanovic, J. J., Rodriguez, N., and Pratt, T. C. (2012). The collateral consequences of incarceration revisited: A qualitative analysis of the effects on caregivers of children of incarcerated parents. Criminology, 50(4), 913-959.

Vytlacil, E. (2002). Independence, monotonicity, and latent index models: An equivalence result. Econometrica, 70(1), 331-341.

Western, Bruce. 2006. Punishment and Inequality in America. New York: Russell Sage Foundation.

Western, Bruce and Christopher Muller. 2013. "Mass Incarceration, Macrosociology, and the Poor." The Annals of the American Academy of Political and Social Science 647: 166–89.

Western, Bruce, Leonard M. Lopoo, and Sara S. McLanahan. 2004. "Incarceration and the Bonds between Parents in Fragile Families." Pp. 21–45 in Imprisoning America: The Social Effects of Mass Incarceration, edited by M. Patillo, D. Weiman, and B. Western. New York: Russell Sage Foundation

Western, B. and Pettit, B.: 2010, Collateral costs: Incarceration's effect on economic mobility, Washington, DC: The Pew Charitable Trusts .

Western, B., 2018. Homeward: Life in the Year After Prison.

Wildeman, Christopher. 2010. "Paternal Incarceration and Children's Physically Aggressive Behaviors: Evidence from the Fragile Families and Child Wellbeing Study." Social Forces, 89(1): 285-309.

Wildeman, Christopher, and Bruce Western. 2010. "Incarceration in Fragile Families." The Future of Children, 20(2): 157-177.

Wildeman, Christopher, Signe Hald Anderson, Hedwig Lee and Kristian Bernt Karlson. 2014. "Parental Incarceration and Child Mortality in Denmark." American Journal of Public Health, 104(3): 428-433.

Tables

Table 1: Population by conviction and incarceration

Sample:	Census: Adult population	SISBEN Criminal record		SISBEN w/ conviction By incarceration	
		No	Yes	No	Yes
	(1)	(2)	(3)	(4)	(5)
Years of education	7.36	6.82	6.68	6.86	6.42
Finished High School D=1	44.0%	31.2%	22.8%	24.2%	20.8%
Income score		34.01	30.90	31.72	29.41
Gender (Male=1)	49.0%	47.6%	83.3%	84.5%	83.3%
# HH members	3.90	4.28	4.47	4.37	4.43
Occupation: Working D=1	48.0%	47.3%	65.4%	67.0%	63.9%
Head of the household D=1		41.2%	47.1%	46.9%	48.6%
Year of birth	1965	1966.9	1974.8	1975.0	1974.3
Marital status: Single D.	45.0%	34.7%	40.7%	45.0%	43.6%
Obs	26,757,687	16,195,178	89,257	55,790	33,467
Years of education for children	8.41	7.20	6.71	6.93	6.57

Notes: Columns 1-5 are group means. HHH: Head of the household, HS: High School. D: Dummy. Income Score: Score from 0 to 100, calculated using variables on income and education of the members of the household, size and characteristics of the house. Source: 2005 Census, SISBEN and criminal records.

Table 2: Convicted parents by incarceration and gender

Convicted sample: by gender and incarceration status	Women		Men	
	No	Yes	No	Yes
	(1)	(2)	(3)	(4)
Years of education	6.50	6.06	6.68	6.23
Dummy Has HS degree =1	20%	16%	22%	19%
Income Score	17.2	16.1	19.48	18.46
Occupation: Dummy Working=1	45%	40%	69%	68%
Dummy head of the household=1	36.2%	37.1%	47%	50%
Age at sentence	35.5	36.2	34.46	36.31
Marital status: Dummy Single=1	47.8%	45.1%	46%	44%
Obs	9,375	6,028	46,415	27,439

Notes: Columns 1-4 are group means. HHH: Head of the household, HS: High School. D: Dummy. Income Score: Score from 0 to 100, calculated using variables on income and education of the members of the household, size and characteristics of the house. Source: SISBEN and criminal records.

Table 3: First stage - Parents

Dep var: Decision Dummy	(1)	(2)	(3)	(4)
	Conviction	Conviction	Incarceration	Incarceration
Judge Stringency	0.697***	0.697***	0.785***	0.786***
	[0.0368]	[0.0368]	[0.0432]	[0.0430]
Controls		X		X
F stat*	4.4	4.0	3.9	4.2
F critical value	4	4	4	4
Obs	233,050	116,062	90,774	90,774
Judges	1395	1369	770	770
R-sq	0.124	0.124	0.241	0.242
adj. R-sq	0.118	0.118	0.235	0.237

Controls column 2: Gender, age, number of crimes, and crime category. The number of observation changes from column 1 to 2 because control variables are added to the data base at the discretion of a randomly assigned prosecutor. Controls column 3: Randomization unit fixed effects and Pc. Column 4 adds: Years of education, gender, income score, year of birth, occupation, year of survey. Standard errors clustered at the randomization unit level. Sources: Attorney General's Office, criminal records and poverty census. F-stat is calculated from a regression on judge dummys.

Table 4: Balance test-Trial sample

Dep. Var: Conviction / Incarceration stringency	Judge: Conviction stringency	Judge: Incarceration stringency
Age	0.0000024	0.00000914
	[0.0000208]	[0.0000354]
Gender	0.000324	-0.000291
	[0.000509]	[0.000753]
Number of charges	0.000867	0.000718
	[0.000835]	[0.00157]
Violent crime	-0.000293	0.0014
	[0.000805]	[0.00129]
Property crime	0.00203	0.00117
	[0.00224]	[0.00360]
Drugs related crime	-0.000927	-0.00189
	[0.00157]	[0.00271]
Guns related crime	-0.000666	-0.00101
	[0.00142]	[0.00213]
Misdemeanor	-0.000867	0.00139
	[0.00112]	[0.00183]
Obs	187,231	162,960
Judges	1,272	683
F-test	0.52	0.80

Standard errors clustered at the randomization unit/year level. Each rows corresponds to a different regression of judge leniency and defendant characteristics. When testing balance across crime categories I construct an alternative measure of conviction stringency that doesn't parse-out crime level conviction rates. The F-test corresponds to a regression where I include all the variables at the same time. Source Attorney General's office and criminal records.

Table 5: Balance test II-Incarcerated sample

Dep var: Incarceration FE	**(1)** 0.74<Pc<0.88	**(2)** 0.88<Pc<0.9	**(3)** 0.9<Pc<1	**(4)** Pooled Pc
Years of education	-0.0000292	-0.0000215	0.000274	0.00011
	[0.000119]	[0.000136]	[0.000169]	[0.0000873]
Income score	-0.0000174	0.00000267	0.000013	0.0000106
	[0.0000283]	[0.0000292]	[0.0000364]	[0.0000175]
Age at sentence	0.0000218	-2.08E-08	0.0000107	0.0000197
	[0.0000338]	[0.0000320]	[0.0000435]	[0.0000266]
Gender	-0.00142	0.001	-0.00212**	-0.00104
	[0.00127]	[0.000793]	[0.00100]	[0.000633]
Years of education HH	-0.0000463	0.000106	-0.000153	-0.0000165
	[0.000157]	[0.000136]	[0.000162]	[0.0000996]
D: Working	-0.0000919	-0.000981	0.000137	-0.000126
	[0.000672]	[0.000763]	[0.00108]	[0.000493]
D: Studying	-0.0022	-0.000602	0.00103	0.00108
	[0.00316]	[0.00278]	[0.00364]	[0.00199]
D: Both census surveys	-0.000844	-0.000942	0.000587	-0.000305
	[0.000897]	[0.000634]	[0.000857]	[0.000488]
D: First survey	0.000355	0.000691	0.000648	0.000511
	[0.00124]	[0.00123]	[0.00162]	[0.000800]
Constant	0.178*	-3.04E-01	6.64E-02	0.360***
	[0.107]	[0.226]	[0.124]	[0.00594]
F Test	0.8494	0.5001	0.564	0.5763
Obs	16,684	17,416	15,845	49,945
R-sq	0.128	0.149	0.137	0.03

Additional controls: Pc, Randomization unit FE, sentence year FE. Standard errors clustered at the randomization unit year level.

Table 6: Monotonicity

Monotonicity test: Out-of-sample First stage

	Males	Females	Violent	Not violent	Young	Old
Conviction-Judge FE	0.789***	0.194***	0.164***	0.376***	0.334***	0.310***
Out of sample	[0.0520]	[0.0102]	[0.00870]	[0.0208]	[0.0278]	[0.0198]
Obs	20,665	147,066	143,567	75,345	50,267	70,042
Incarceration-Judge FE	0.587***	0.163***	0.0517***	0.189***	0.360***	0.451***
Out of sample	[0.0565]	[0.0148]	[0.0163]	[0.0275]	[0.0237]	[0.0336]
Obs	23,345	104,672	78,652	48,582	75,710	50,387

Table 7: Monotonicity test: Norris

Pairwise Monotonicity Test	P-value
Gender	0.33
Primary school	0.99
Young	0.93
Single	0.99
Poor	0.99
Working	0.86
Type of crime:	
Violent	0.52
Property	0.00
Gun-related	0.38
Drug-related	0.32

Norris (2019) test for monotonicity.

Table 8: Monotonicity Test: Frandsen et al

Randomization Unit	**Critical value**	**P-value**
1	28.561	0.435
2	36.685	0.302
3	22.108	0.279
4	11.612	0.071
5	0.698	0.983
6	5.372	0.372
7	1.197	0.754
8	10.637	0.014
9	2.362	0.501
10	4.485	0.214
11	0.465	0.495
12	0.997	0.607
13	0.265	0.876
14	0.007	0.931
15	4.083	0.130
16	3.72	0.156
Joint test	133.254	0.160

Frandsen et al (2019) test for Monotonicty. I run the test in the randomization units were there are more than 4 judges which corresponds to 73% of my sample.

Table 9: OLS Regression

	Children with a convicted parent by age 14			
OLS: no controls	(1)	(2)	(3)	(4)
Dep var: Years of education	Pooled Pc	0.7<Pc<0.88	0.88<Pc<0.9	0.9<Pc<1
Parental Incarceration Dummy	**-0.356*** **	**-0.397*** **	**-0.319*** **	**-0.372*** **
	[0.0717]	[0.0896]	[0.0815]	[0.0894]
Constant	6.768***	6.446***	6.586***	6.693***
	[0.0976]	[0.0857]	[0.0806]	[0.101]
OLS: Adding controls				
Parental Incarceration Dummy	-0.0764***	-0.0529	-0.104**	-0.0608*
	[0.0235]	[0.0426]	[0.0470]	[0.0346]
Obs	52,275	16,091	16,981	16,424
Clusters	609	329	365	403
R-sq	0.383	0.406	0.375	0.364

Controls: Randomization unit FE, Gender, YOB FE, Sisben score, gender of incarcerated parent, pc, year of sentence, birth order and year of survey. Sample: Children between 1990 and 2007 who had a convicted parent between ages 0 and 14. SE in brackets clustered at the randomizarion unit. AR confidence interval result in the same significance levels.

Table 10: Results: Reduced form and IV

Reduced form	(1)	(2)	(3)	(4)
Dep var: Years of education	Pooled Pc	0.7<Pc<0.88	0.88<Pc<0.9	0.9<Pc<1
Judge leave-out incarceration rate	**0.763*** **	**0.720***	**0.901***	**0.741** **
	[0.215]	**[0.419]**	**[0.503]**	**[0.341]**
R-sq	0.298	0.322	0.286	0.292
IV Dep var: Years of education	Pooled Pc	0.7<Pc<0.88	0.88<Pc<0.9	0.9<Pc<1
Parental Incarceration Dummy	**0.670*** **	**0.670***	**0.940***	**0.633** **
	[0.194]	**[0.401]**	**[0.552]**	**[0.286]**
Obs	51,742	16,086	16,979	16,416
Clusters	603	324	363	395

Controls: Randomization unit FE, Gender, YOB FE, Sisben score, gender of incarcerated parent, pc, year of sentence, birth order and year of survey. Column 1 controls add a second order polynomial on Pc. Sample: Children between 1990 and 2007 who had a convicted parent between ages 0 and 14. SE in brackets clustered at the randomizarion unit. AR confidence interval result in the same significance levels.

Table 11: Heterogeneous effects

IV	**Girls**	**Boys**	**Mother**	**Father**
Dep var: Years of education	(1)	(2)	(3)	(4)
Parental Inc.	0.359*	0.865***	0.823**	0.531**
	[0.208]	[0.286]	[0.370]	[0.222]
Obs	26310	27086	12049	41319
	Type of crime			
	Violent	Property	Drug-related	Gun-related
Parental Inc.	1.634	0.485	0.734**	0.542
	[1.167]	[0.606]	[0.361]	[0.508]
Obs	9792	12985	12905	9857
Pooled Pc	x	x	x	x

Controls: Gender, YOB FE, Sisben score, gender of incarcerated parent, pc, year of sentence, birth order and year of survey. Sample: Children between 1990 and 2007 who had a convicted parent between ages 0 and 14. SE in brackets clustered at the randomizarion unit.

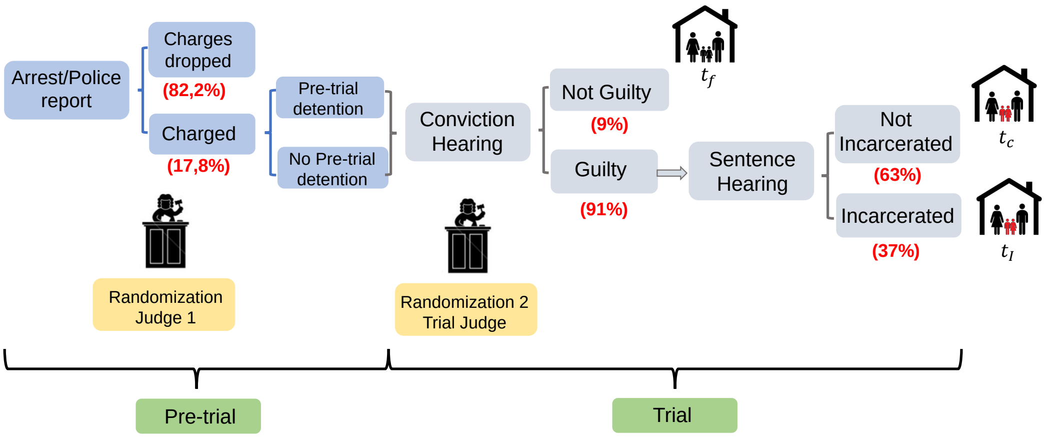
Table 12: Changes after incarceration

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep var:	LFP spouse	Income score	Years of educ. HHH	D: Male HHH	# of people in HH	D: Lives w/ Grandparents	D: In 2nd SISBEN
Parental Inc.	0.0680***	-2.365***	0.103***	-0.0786***	-0.0996***	0.0196*	-0.0303***
	[0.0187]	[0.193]	[0.0300]	[0.00604]	[0.0303]	[0.0110]	[0.00492]
Obs	9,673	82,779	82,779	82,779	81,615	16,372	32,881
R-sq	0.22	0.75	0.20	0.19	0.33	0.10	0.08
Mean dep var:	0.399	26.41	5.1	0.595	4.659	0.215	0.242
St dev dep var:	0.49	20.13	2.911	0.491	2.42	0.411	0.428

Panel regressions. Controls: Poverty score, years of education of HHH, Municipality FE and year of survey FE. Dummy for living with grandparents also includes uncles and cousins. Households with data on both cross-sections of the poverty census and who had a conviction episode in between surveys. Source: SISBEN and criminal records.

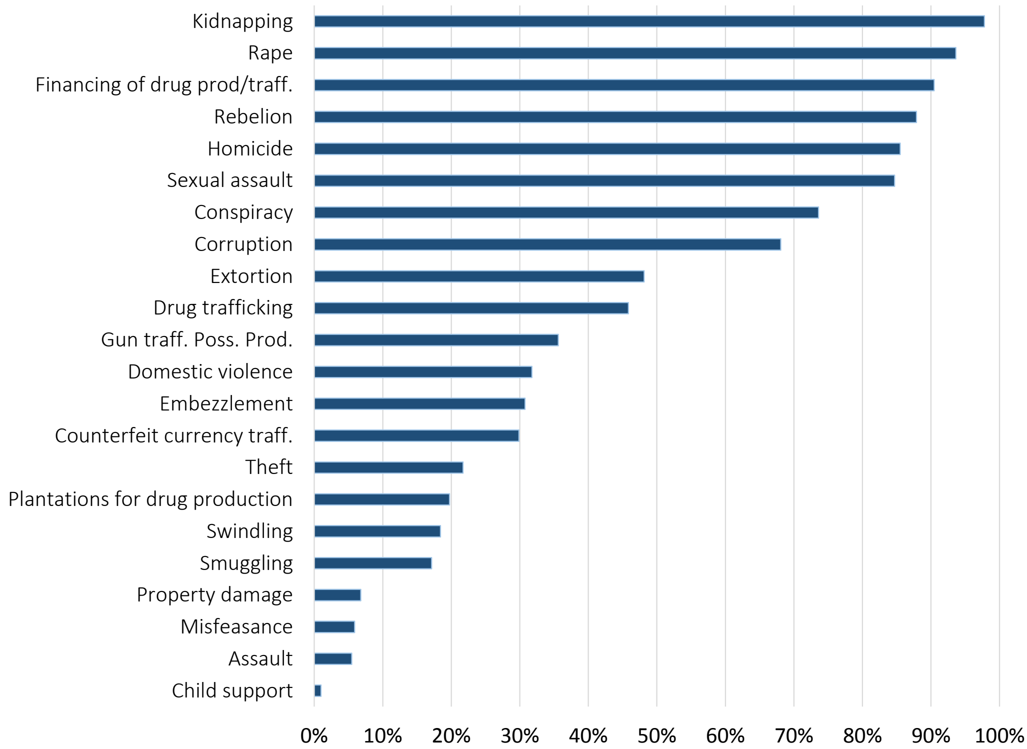
Figures

Figure 1: Prosecution and trial stages



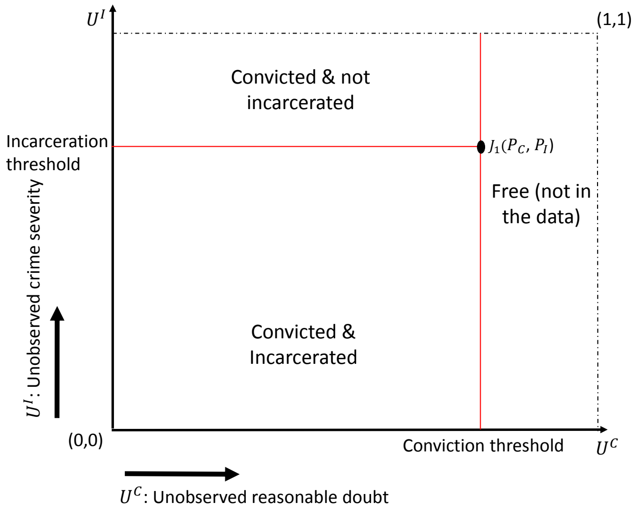
Source: Colombian Penal proceedings code, Informe de la Comision Asesora de Politica Criminal (2012), SPOA and Criminal records. The treatment status studied in this paper corresponds to t_f, which refers to parents who are not convicted or free, t_c those convicted but not incarcerated, and t_I those convicted and incarcerated. Incarceration is a function of sentence length. Currently, a sentence equal to four years or less is not served in prison.

Figure 2: Incarceration rates



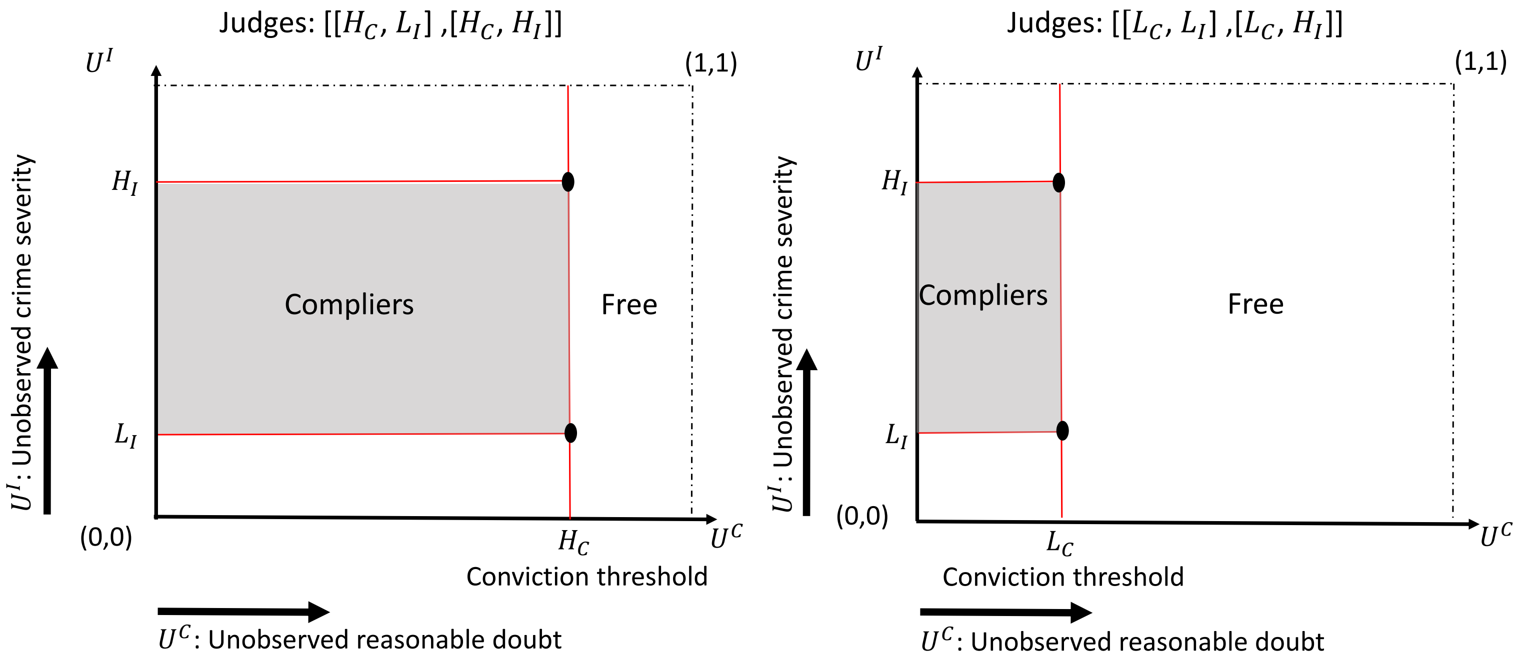
Source: Criminal records. Selected crimes. I restrict to crimes with at least 100 cases.

Figure 3: Identification: Defendant's space, judges thresholds and treatment assignment



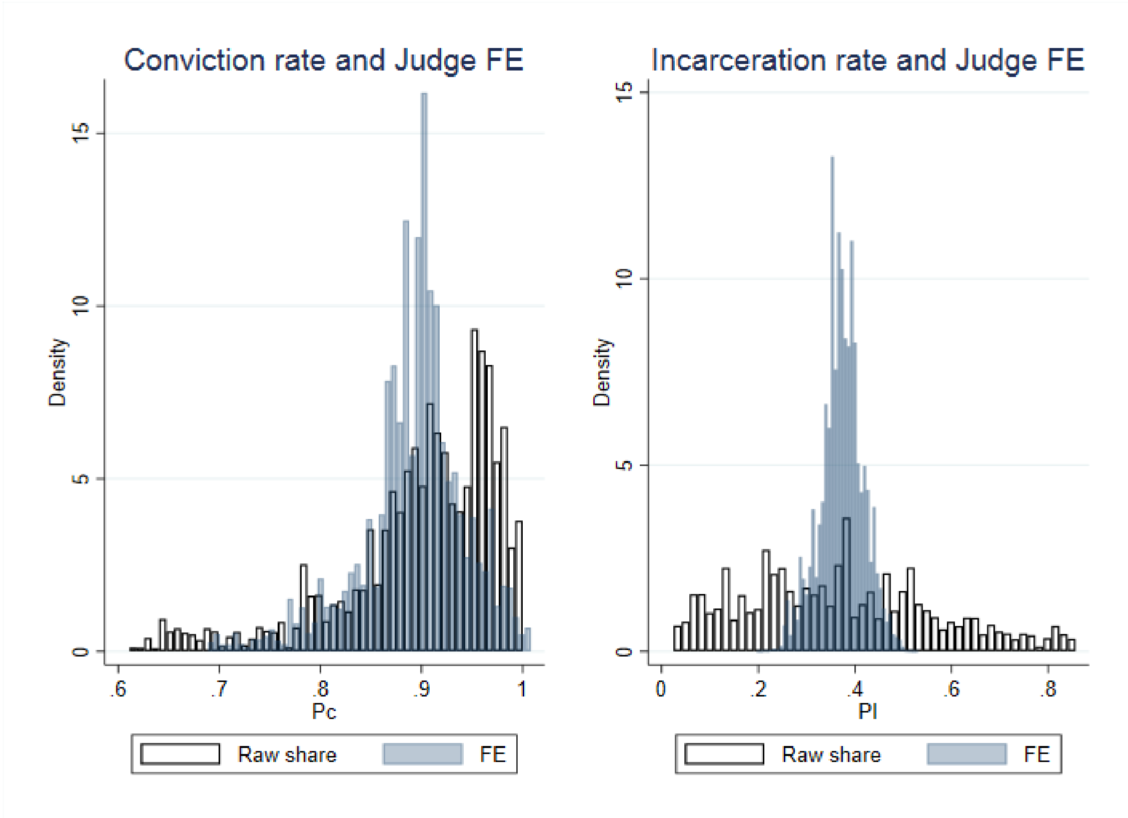
A defendant is characterized by a point in the unitary square. A judge is defined by a pair of thresholds along each axis which determine treatment assignments. Defendants to the left of the conviction threshold are convicted, and those to the right are freed. Among the convicted, defendants below the incarceration threshold go to prison, and those above do not.

Figure 4: Identification under 4 types of judges



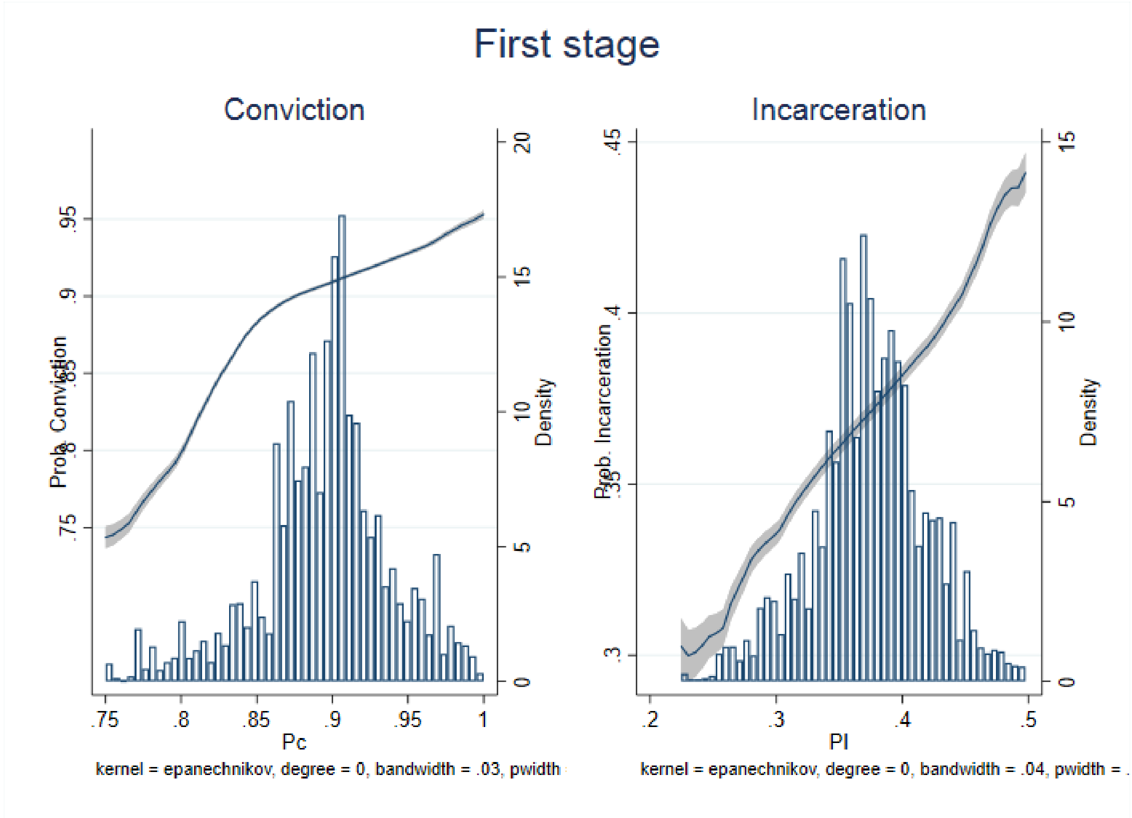
The left panel features harsh judges on the conviction margin (H_c). This judges can be harsh (H_I) or lenient (L_I) on the incarceration margin. We can identify the causal effect of incarceration for defendants in the shaded area. Those whose incarceration decision is only a function of judge assignment. The right panel is analogous and it features lenient judges on the conviction margin (L_c).

Figure 5: Judges' fixed effects



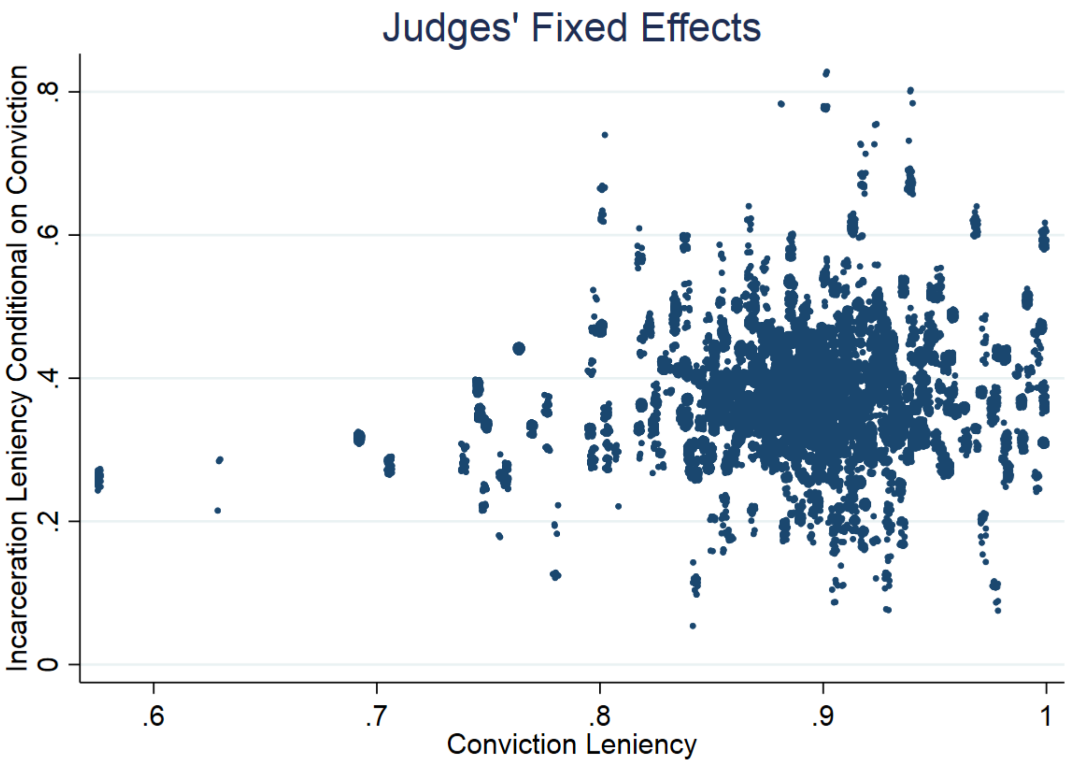
Source: Attorney General's office and criminal records. Raw rates are conviction/incarceration averages by judge. To construct the judge's fixed effect I take the residual at the judge level after regressing conviction/incarceration on (demeaned) randomization unit/year dummies, (demeaned) crime-level conviction/incarceration rates, without a constant.

Figure 6: First stage



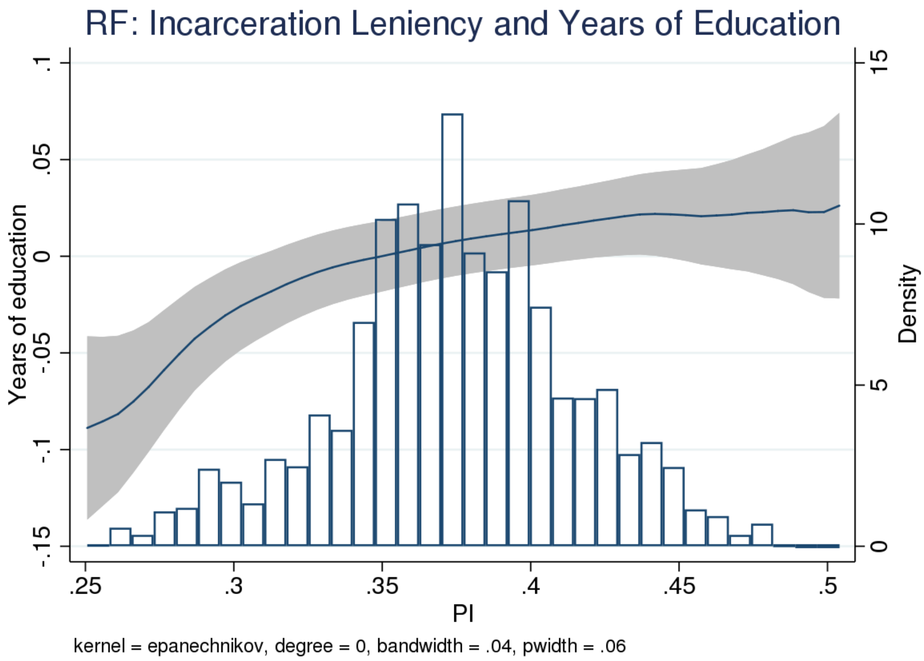
Source: Attorney General's office and criminal records. Raw rates are conviction/incarceration averages by judge. To construct the judge's fixed effect I take the residual at the judge level after regressing conviction/incarceration on (demeaned) randomization unit/year dummys, (demeaned) crime-level conviction/incarceration rates, without a constant.

Figure 7: Scatter plot: Judges' fixed effects



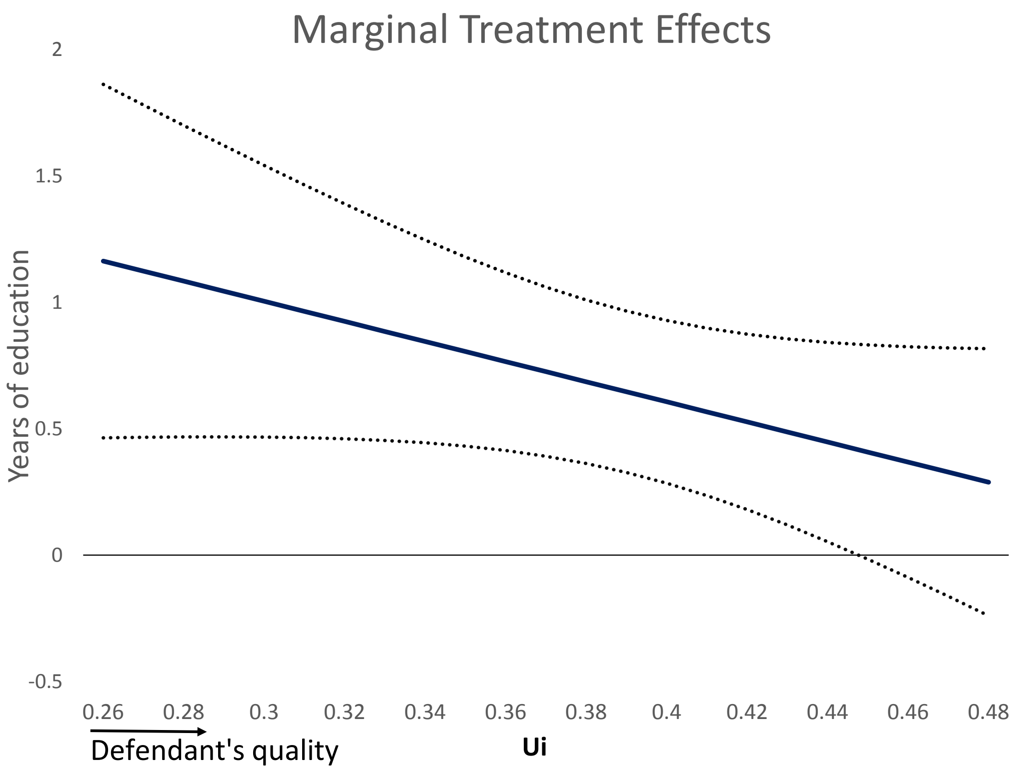
Source: Attorney General's office and criminal records. To construct the judge's fixed effect I take the residual at the judge level after regressing conviction/incarceration on (demeaned) randomization unit/year dummies, (demeaned) crime-level conviction/incarceration rates, without a constant.

Figure 8: Reduced form



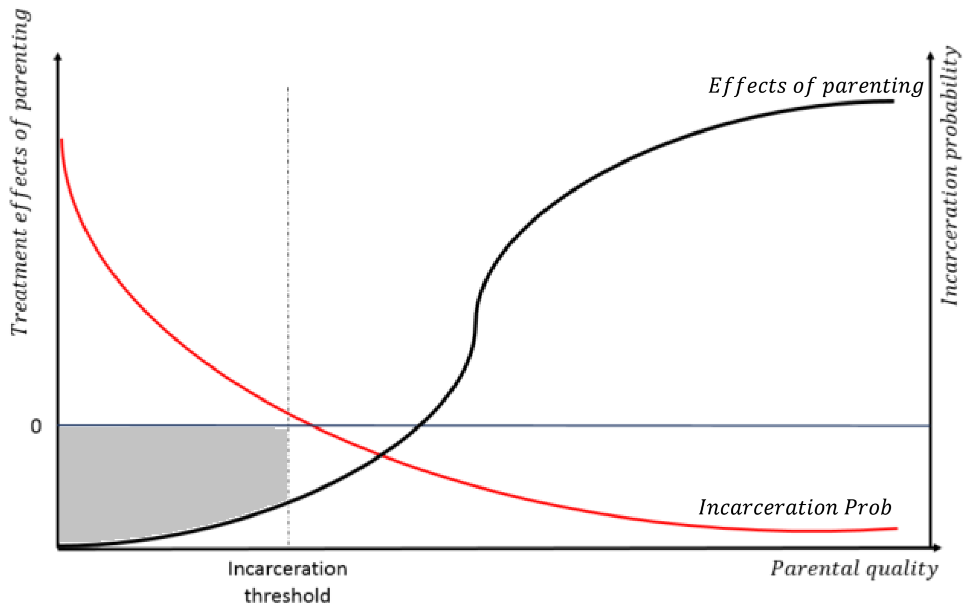
Notes: Histogram of parental incarceration judge leniency and the fitted value of local polynomial regressions of children's educational attainment on judge stringency. Pooled regression I control for p_c.

Figure 9: MTE



Notes: Following the LIV approach in Heckman and Vytlacil (2005) I regress $Yeduc = \alpha + \beta_1 P_i + \beta_2 P_i^2 + \beta_3 X$. This graphs plots: $\beta_1 + 2\beta_2 P_i$ for the pooled regression. Controls: Municipality FE, gender, YOB FE, Sisben score, years of education HH head, years of education of incarecerated parent, gender of incarcerated parent, pc, year of sentence, birth order and year of survey.

Figure 10: Model of parenting and incarceration



A Appendix: Model and proofs

This section continues with the discussion of section 4.2. For ease of exposition, I will first explore identification under the assumption that $U^c \perp U^I$ and then I will go over the results without it. Under the independence assumption we can identify $P_I(z)$ from the data, that is:

$$P(U^I < P_I(z)|U^c \leq P_c(z)) = P(U^I < P_I(z)) = P_I(Z)$$

The left hand side is observed from the data, the first equality follows directly from the independence assumption and the last one the uniform distribution of U^I. P_I is interpreted as the share incarcerated.

The goal is to identify and evaluate the treatment effect: $E(Y(t_I) - Y(t_c))$ which is a function of counterfactual variables $Y(t_I)$ and $Y(t_c)$. To achieve this goal, it is useful to express the observed expectations in terms of the variables that define the model:

$$E(Y \cdot \mathbf{1}[T = t_c]|P_c(Z) = p_c, P_I(Z) = p_I) = \tag{14}$$

$$= E(Y(t_c) \cdot \mathbf{1}[T = t_c]|P_c(Z) = p_c, P_I(Z) = p_I) \tag{15}$$

$$= E(Y(t_c) \cdot \mathbf{1}[U^c \leq p_c] \cdot \mathbf{1}[U^I > p_I]|P_c(Z) = p_c, P_I(Z) = p_I) \tag{16}$$

$$= E(Y(t_c) \cdot \mathbf{1}[U^c \leq p_c] \cdot \mathbf{1}[U^I > p_I]) \tag{17}$$

$$= \int_0^{p_c} \int_{p_I}^1 E(Y(t_c)|U^c = u^c, U^I = u^I) f_{u^c u^I}(u^c, u^I)\, du^c\, du^I \tag{18}$$

$$\tag{19}$$

$$= -\int_0^{p_c} \int_0^{p_I} E(Y(t_c)|U^c = u^c, U^I = u^I) f_{u^c,u^I}(u^c, u^I)\, du^c\, du^I + \int_0^{p_c} E(Y(t_c)|U^c = u^c) f_{u^c}(u^c)\, du^c$$

Equation (14) is an expectation observed in the data. Equality (15) comes from the definition of observed outcomes. Equality (16) expresses the indicator $\mathbf{1}[T = t_c]$ in terms of the inequalities of the choice model. Equality (17) uses the independence relation $Z \perp (U^c, U^I)$. Equality (18) expresses the expectation as the integral over the distribution of U^c, U^I where $f_{U^c,U^I}(u^c, u^I)$ stands for the probability density function of U^c, U^I at the point (u^c, u^I), and is equal to one. Equality (19) modifies the integration region. This change is useful to apply the Lebesgue differentiation theorem next;

$$\frac{\partial^2 E(Y \cdot \mathbf{1}[T = t_c]|P_c(Z) = p_c, P_I(Z) = p_I)}{\partial p_c \partial p_I} = -E(Y(t_c)|U^c = p_c, U^I = p_I) \tag{20}$$

Equality (20) arises as a direct application of the Lebesgue differentiation theorem. What this result gives me is a connection between the observed outcomes and the targeted counterfactual outcome. We can use the same steps applied to counterfactual $Y(t_c)$ to obtain the counterfactual for $Y(t_I)$. Combining these two I obtain:

$$\frac{\partial^2 E(Y \cdot \mathbf{1}[T \in \{t_c, t_I\}]|P_c(Z) = p_c, P_I(Z) = p_I)}{\partial p_c \partial p_I} = E(Y(t_I) - Y(t_c)|U^c = p_c, U^I = p_I) \tag{21}$$

In the language of Heckman and Vytlacil (2005), Eq.21 defines the marginal treatment effect (MTE) of outcome Y with respect to treatment assignment t_c and t_I. It is interpreted as the causal effect of incarceration versus conviction only, for the share of defendants whose culpability and punishment assessments, U^c and U^I respectively, is set at quantiles p_c and p_I. The derivative in Equation (19) traces the MTE of incarceration relative to conviction throughout the unitary square of U^c, U^I. This result is an application of Lee and Salanie (2018) and extends the result of Heckman and Vytlacil (1999). In Appendix B I explain graphically the intuition of this result. The main idea is that changes in P_c and P_I affect exogenously treatment assignment. Then, by examining the derivative of the outcome variables with respect to P_c and P_I, we capture how the outcome variable changes when treatment changes at each point in the space of the unobservable confounding variables.

The average treatment effect (ATE) is the causal effect of t_c and t_I on Y in the population, and it corresponds to the integral of the MTE over the support of U^c and U^I.

$$E(Y(t_I) - Y(t_c)) = \int_0^1 \int_0^1 \frac{\partial^2 E(Y \cdot \mathbf{1}[T \in \{t_c, t_I\}]|P_c(Z) = p_c, P_I(Z) = p_I)}{\partial p_c \partial p_I} dp_c dp_I \tag{22}$$

Without the assumption of independence of U^c and U^I, variation in P_I is only identified once I fix the conviction threshold. Thus, the counterfactual of interest is now: $Y(t_I)$ and $Y(t_c)$ for those who were convicted under $P_c = p_c$. This means the objective is to identify causal effects of the form: $E(Y(t_I) - Y(t_c)|U^c < p_c)$, which is the the same exercise explained in Section 4.1. Let:

$$E(Y \cdot \mathbf{1}[T = t_c]|P_c(Z) = p_c, P_I(Z) = p_I, U^c < p_c) = \tag{23}$$

$$= E(Y(t_c) \cdot \mathbf{1}[T = t_c]|P_c(Z) = p_c, P_I(Z) = p_I, U^c < p_c) \tag{24}$$

$$= E(Y(t_c) \cdot \mathbf{1}[U^I > p_I]|P_c(Z) = p_c, P_I(Z) = p_I, U^c < p_c) \tag{25}$$

$$= E(Y(t_c) \cdot \mathbf{1}[U^I > p_I]|U^c < p_c) \tag{26}$$

Where I followed the same steps as before. Let:

$$P_I^* = Pr[U^I < P_I|U^c < P_c] = G(P_I) \tag{27}$$

P_I^* is the object I observe so I will define the observed expectations in terms of this variable:[33]

$$E(Y(t_c) \cdot \mathbf{1}[U^I > G^{-1}(p_I^*|U^c < p_c]|U^c < p_c) \tag{28}$$

$$\int_{P_I^*}^1 E(Y(t_c)|U^I = u^I, U^c < p_c) f_{u^{I*}|U^c<p_c}(p_I^*) du^I \tag{29}$$

[33]Where $f_{u^{I*}|U^c<p_c}(p_I^*)$ in eq. (39) corresponds to:$f_{u^I|U^c<p_c}(p_I)\frac{\partial P_I((p_I^*)}{(p_I^*}$

And applying the Lebesgue differentiation theorem this results in:

$$\frac{\partial E(Y\cdot \mathbf{1}[T\in\{t_c\}]|p_c,p_I,U^c<p_c)}{\partial p_{I^*}} = -E(Y(t_c)|U^I=p_I,U^c<p_c)f_{u^I|U^c<p_c}(p_I^*) \quad (30)$$

And ultimately;

$$E(Y(t_I)-Y(t_c)|U^c<p_c) = \int_0^1 \frac{\partial E(Y\cdot \mathbf{1}[T\in\{t_c,t_I\}]|P_c(Z)=p_c,P_I^*(Z)=p_I^*,U^c<p_c)}{\partial p_I^*}dp_I^* \quad (31)$$

What this result says is that we can trace the treatment effect of incarceration relative to conviction once we fix a threshold for conviction. We do this by evaluating the changes on the outcome variable when we change P_I^*. This delivers the MTE along the unobservable dimension $U^I|U^c<P_c$. The integral over the support of the instrument gives the LATE, or the ATE when the instrument has full support.

B Appendix: Intuition for the 2 dimension LATE

In this section I go over the intuition of the results in eq. 18 and eq.19. This result extends the intuition behind LATE to a two-dimensional space. To make this point clear, let us think in discrete terms and use an example with 4 judges with threshold levels $\{P_c^1,P_I^1\}$, $\{P_c^1,P_I^2\}$, $\{P_c^2,P_I^1\}$, and $\{P_c^2,P_I^2\}$.[34]

For notation purposes, let:

$$f(p_c,p_I) = E(Y\mathbf{1}[T\in\{t_c\}]|P_c(Z)=p_c,P_I(Z)=p_I) \quad (32)$$

and

$$g(p_c,p_I) = E(Y\mathbf{1}[T\in\{t_I\}]|P_c(Z)=p_c,P_I(Z)=p_I) \quad (33)$$

Next, I can rewrite, in discrete terms, the identification result in equation 5 as:

$$\frac{\Delta f(p_c,p_I)}{\Delta p_c\Delta p_I}+\frac{\Delta g(p_c,p_I)}{\Delta p_c\Delta p_I} = [f(p_c^2,p_I^2)-f(p_c^1,p_I^2)]-[f(p_c^2,p_I^1)-f(p_c^1,p_I^1)]+ [g(p_c^2,p_I^2)-g(p_c^1,p_I^2)]-[g(p_c^2,p_I^1)-g(p_c^1,p_I^1)] = E(Y(t_I)-Y(t_c)|u^c=p_c,u^I=p_I) \quad (34)$$

Now, let us go over each term in (31). First, $f(p_c^2,p_I^2)$ represents the outcomes of convicted but not incarcerated individuals who had a judge with thresholds $\{P_c^2,P_I^2\}$. Panel a in Figure C.3 shades the area in the u^c, u^I square that identifies these individuals. The next panels in Figure C.4 highlight the following terms in equation 8 and their differences. Ultimately, what equation (31)

[34] Equivalent to {HL}, {HH}, {LH}, and {LL} in Section 4.

is doing is identifying the complier range in a two-dimensional space, which instead of an interval is a rectangle.

I estimate (18) by fitting a polynomial on p_I and p_c and evaluating the cross-derivative on the support of the instruments. Figure C5 shows the MTE in the relevant segment of the (u^c, u^I) square. There are some interesting features of these results; first, as before, as we increase u^I (defendants' quality), the effect on years of schooling decreases, confirming that this positive effect is accrued when incarceration removes a bad parent from the household. What is new in Figure C.5 is that now we can also move along the u^c margin, or the "strength of the evidence" margin. The data also show that as evidence becomes weaker, the positive effects also decrease. Ultimately, what this exercise shows is that the effect on children is very sensitive to the type of case a judge is deciding on. In the case of Colombia, marginal incarcerations are of defendants still very negatively selected and with sufficient evidence against them, so that their children are better off without that parent. How this result extends to other settings is a function of the location of the marginal cases in the u^c, u^I square.

C Appendix : Data construction

In this appendix, I explain in detail the construction of the sample and variables I use throughout the paper. The starting point for my data construction are the two SISBEN surveys. These data are collected by the government to target social programs for the poor. The survey is conducted at the household level, and consists of two modules. In the first, it asks about the characteristics of the house (flooring material, number of bedrooms, etc), access to utilities, and assets in the households (TV, refrigerator, car, etc.). In the second part, all members of the household are listed with names and national identification numbers, and their relationship to the head of the household is specified. The questionnaire then asks about gender, age, education level, marital status, disability status, and occupation. This survey is applied to everyone living in a municipality with a population of 30,000 or less, and in larger municipalities local authorities target households who could be potential beneficiaries of welfare programs. If a household is not targeted by local authorities and wishes to be surveyed, it can easily request to be included. The government uses this information to create a formula that measures the household's ability to provide resources for its members, and computes a score for each household that determines eligibility for different social programs. These data provide me with i) identification numbers with municipality location to web-scrape criminal records and, ii) parent-to-child links.

I select the population of adults who lived in the 17 out of 33 municipalities that have criminal records online. These districts represent 67% of the population, and 69% of homicide and 83% of property crimes.[35] I then web-scrape criminal records (from http://procesos.ramajudicial.gov.co/consultaprocesos/) by selecting the district and then searching individually for records with the ID numbers. From a 5% sample in which I look for criminal records in all 17 districts I estimate that I will miss 8.6% of the sample due to crimes committed in districts different from the one in the SISBEN.

I find 328,937 criminal records that belong to 256,366 individuals. I start by dropping observations that have missing values in year of sentence, crime or courtroom identifier (81,049 observations

[35] Judicial districts with online data: Armenia, Barranquilla, Bogota, Bucaramanga, Buga, Cali, Ibague, Florencia, Manizales, Medellin, Neiva, Palmira, Pasto, Pereira, Popayan, Tunja, and Villavicencio.

deleted). Next, I drop all records before 2005 and records earlier to the year of the first SISBEN records(59,872 observations deleted), and all cases in which there is only one judge per district (4,635 observations deleted). I keep only the courtrooms for which there is data on convictions (14,786 observations deleted). Finally, I drop all observations where there are less than 15 cases in a year/judge cell (56,268 observations deleted). After this, I end up with 112,696 criminal records which correspond to 93,676 individuals. Table C.1 shows differences between the characteristics of individuals in the final data-set and those who were dropped. For the set of observations that have sentence data, I find that there is no evidence of differential incarceration rates across samples.

To assess how representative my sample is of the prison population, I compare counts of individuals sentenced by year from my data with counts of new inmates from official records of the Prison Authority (INPEC). I only have information available for 2015; according to INPEC, there were 27,287 new immates that year, from my data, I find that 5,932 defendants were sent to prison, which would suggest that I have data on 22% of the prison population. This number, however, should be taken with caution, because INPEC data include flows of inmates across prisons, and I don't have data on the size of these flows.

Next, I link these convicts to the 436,309 individuals living in their households, of whom 179,699 are in the relevant cohort years (1991-2007), and 106,465 are the child of a convict. Of this, 96,383 experienced the sentencing episode between ages 0 and 14. Finally, I have education data for 77% of these children. This rate is close to the share of children between ages 12 and 17 who attend school, according to the census (76%). Table C.2 shows regressions of missing education record on parental incarceration. The OLS estimates are negative but very close to zero, once I instrument for incarceration the estimate become more negative but statistically equal to zero. This may suggest that there is also a positive effect on the extensive margin of educational attainment. Missing values are related to the child's not being at school, and for household with lower income and lower education of the head of the household.

Table C1: Sample selection-Defendants

Dep var: Out of sample D.	(1)	(2)
Incarceration		0.00141
		[0.00204]
Years edu.	0.0018	0.00118
	[0.00150]	[0.00157]
Income score	0.00118***	0.000837***
	[0.0000822]	[0.0000879]
Male D.	-0.0400***	-0.0209***
	[0.00279]	[0.00290]
Head HH D.	0.00877**	0.00771**
	[0.00370]	[0.00389]
Single	-0.0298***	-0.0213***
	[0.00222]	[0.00239]
Years edu. HHH	0.0004	0.000919
	[0.00150]	[0.00157]
D: Studying	0.0264***	-0.00653
	[0.00490]	[0.00486]
D: Working	0.0177***	0.0154***
	[0.00209]	[0.00226]
Yob	-0.00708***	-0.00312***
	[0.0000877]	[0.0000956]
Constant	14.55***	6.55E+00
	[0.173]	[3279.3]
Obs	260,968	196,314
R-sq	0.14	0.306

Additional controls: Municipality FE and survey year FE. The first column includes all criminal records and the second restricts to the ones that have data on sentence length.

Table C2: Sample selection

Dep var: Missing Education records.	(1)	(2)	(3)	(4)
Parental incarceration	-0.00125	-0.00191	-0.0615	-0.0597
	[0.00333]	[0.00334]	[0.0548]	[0.0548]
Gender	0.00500*	0.00434	0.00463	0.00395
	[0.00294]	[0.00297]	[0.00293]	[0.00294]
Gender of the parent	-0.0172***	-0.0139***	-0.0166***	-0.0135***
	[0.00391]	[0.00384]	[0.00419]	[0.00402]
Sisben Income score		-0.000408***		-0.00019
		[0.000154]		[0.000177]
Years of education Head of household		-0.00267***		-0.00339***
		[0.000721]		[0.000763]
D: Studying		-0.0709***		-0.0748***
		[0.00566]		[0.00615]
	OLS	OLS	IV	IV
Obs	96383	96383	96380	96380
R-sq	0.133	0.136	0.103	0.107

Additional controls: Year of birth dummys, survey year, birth order, Municipality FE.

Table D1: Sentencing guidelines

Sentencing guidelines	Prison time	
Crime	Colombia	US NY
Possession of cocaine: 14 grams -100 grams	5 to 9 years	1 to 9 years
Assault		
Simple/third degree	1 to 3 years	Up to 1 year
2nd degree	2 to 7 years	3 to 7 years
Theft		
Simple	2 to 9 years	Up to 1 year
Aggravated theft	6 to 14 years	2-7 years
Domestic violence	4 to 8 years	Less than a year to 25 years

Source: Colombia articles 376, 112 239, 240 of the penal code, respectively. For New York: 220.16, 120.00, 120.00, 155.25 or 165.40, 155.30 and 120.00 to 120.12 sections of New York penal law code, respectively.

Table D2: Placebo check

Placebo test			
Dep var: Years of education	OLS	RF	IV
Parental inc.	-0.0182***		0.0609
	[0.00705]		[0.187]
Judge leniency		0.0533	
		[0.143]	
Constant	4.075	3.908	4.152
	[4.106]	[4.103]	[4.085]
Obs	46,257	46,257	46,257

Controls: Randomization unit FE, gender, YOB FE, SISBEN score, years of education head, years of education incarcerated parent, gender of incarcerated parent, Pc, year of sentence, birth order and year of survey. Sample: Children between 1990 and 2007 who had a convicted parent between ages 0 and 14. SE in brackets clustered at the randomization unit and year level.

Table D3: LIML estimates

IV LIML Dep var: Years of education	(1) 0.7<Pc<0.88	(2) 0.88<Pc<0.9	(3) 0.9<Pc<1	(4) Pooled Pc
Parental Incarceration	**0.741****	**0.89**	**0.748****	**0.827*****
	[0.371]	[0.834]	[0.356]	[0.280]
Obs	17,347	18,672	17,045	53,064

Controls: Randomization unit FE, gender, YOB FE, SISBEN score, years of education head, years of education incarcerated parent, gender of incarcerated parent, Pc, year of sentence, birth order and year of survey. Column 4 controls add a second order polynomial on Pc. Sample: Children between 1990 and 2007 who had a convicted parent between ages 0 and 14. SE in brackets clustered at the randomization unit and year level.

D Appendix: Extra tables and figures

Figure D1: Treatment Effects by grade

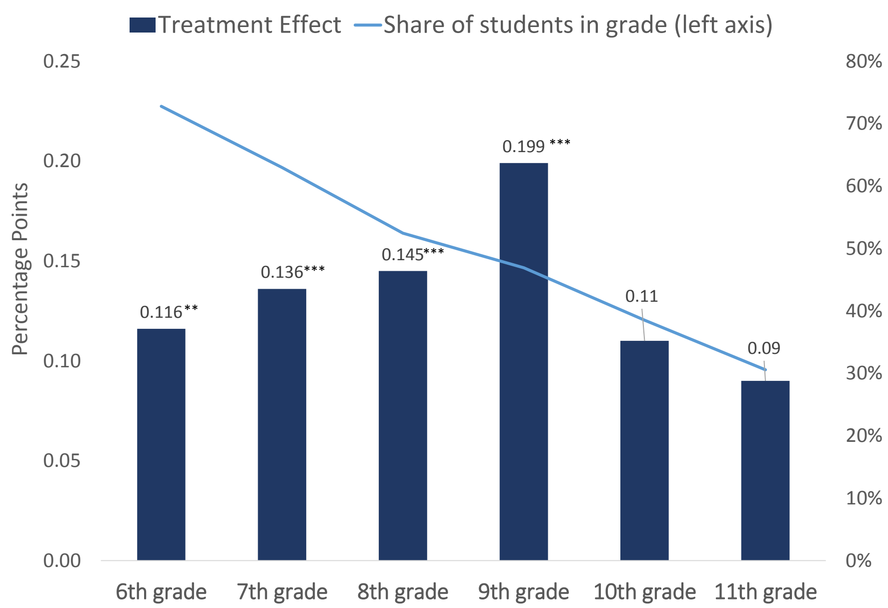
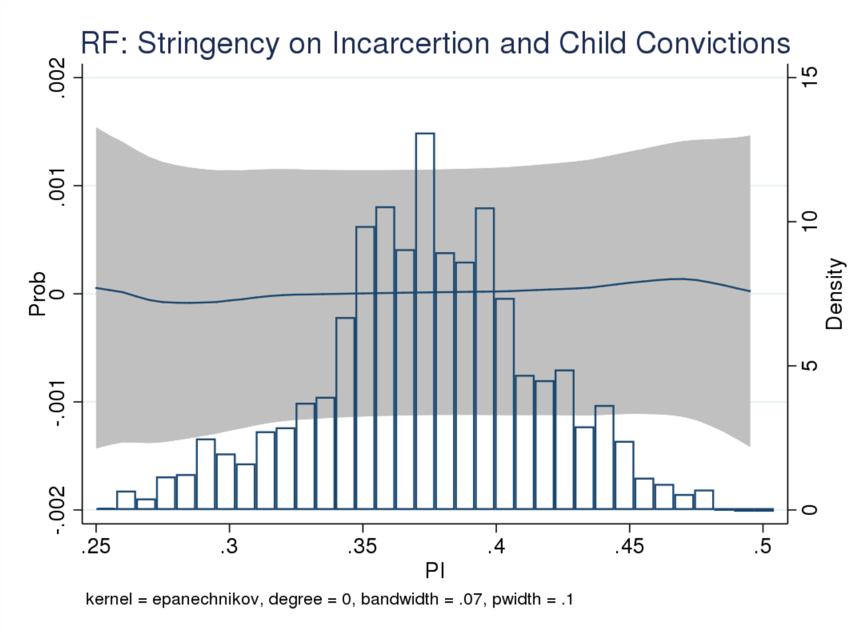
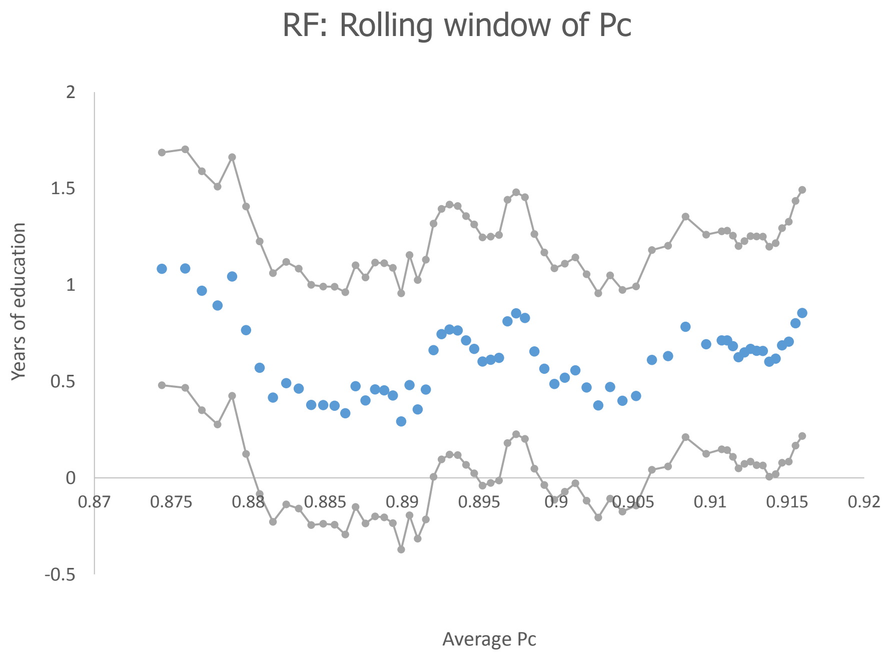


Figure D2: Reduced form



Notes: Histograms of parental incarceration judge stringency and the fitted value of local polynomial regressions of children's criminal records on judge stringency.

Figure D3: Rolling reduced form



Notes: Reduced form estimates of a sample size of 18.000, with a rolling window of 500 on P_c. Grey lines represent 90% confidence intervals.

Figure D4: Identification in 2 dimensions

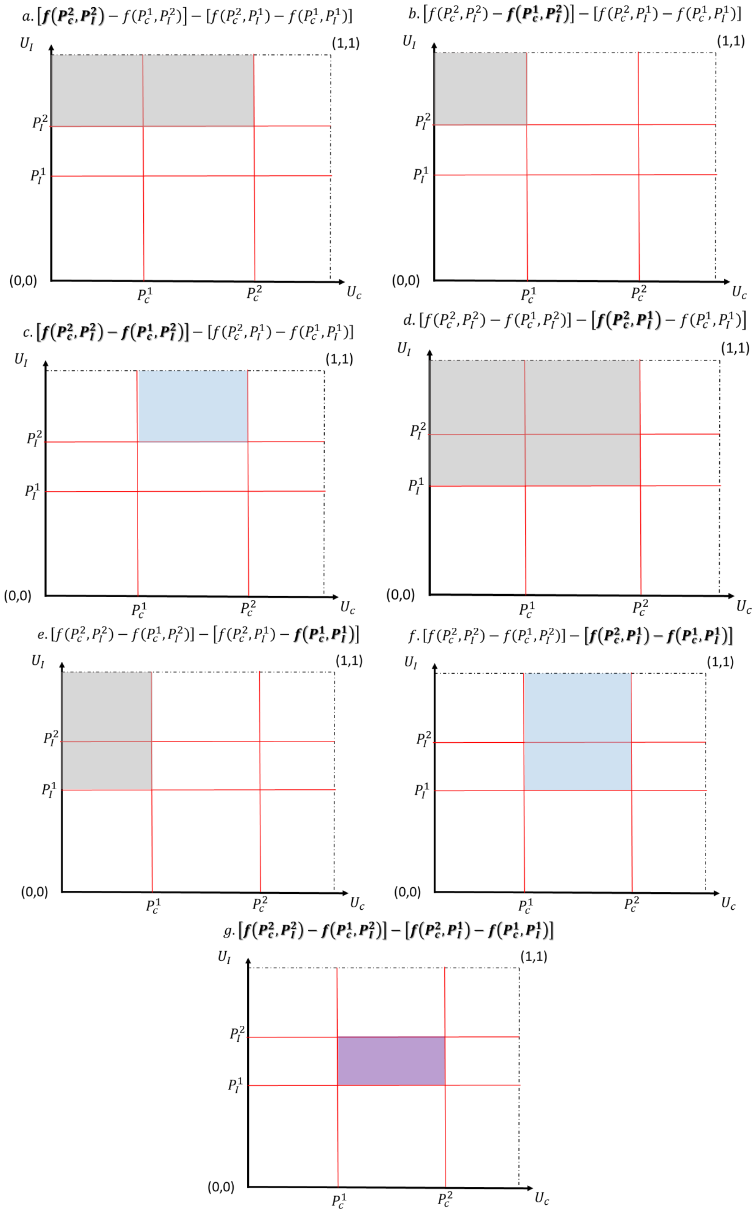


Figure D5: Compliers rectangle

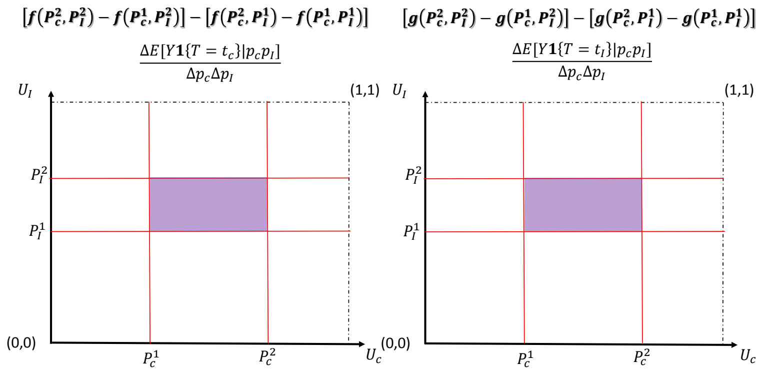


Figure D6: Unconditional MTE